UNIVERSITY OF BIRMINGHAM

INSTITUTE OF GRAVITATIONAL WAVE ASTRONOMY

Astrophysical inference from pulsar timing array searches for gravitational waves

Hannah Rose MIDDLETON

Supervisor:

Alberto VECCHIO

Second supervisor:

Alberto SESANA

Institute of Gravitational Wave Astronomy
Astrophysics and Space Research Group
School of Physics and Astronomy
University of Birmingham
Birmingham, B15 2TT

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

September 2017

Abstract

Gravitational waves (GWs) have been detected for the first time in 2015 by the LIGO-Virgo Scientific Collaboration. The source of the GWs was a binary black hole (BBH). The observation caught the final fraction of a second as the two black holes spiralled together and merged. This observation (and the others to follow) marked the beginnings of GW astronomy, 'a new window on the dark universe', providing a means to observe astronomical phenomena which may be completely inaccessible via other avenues as well as a new testing ground for Einstein's theory of general relativity (GR). However, this is just the beginning – like electromagnetic astrophysics, there is a full spectrum of GW frequencies to explore.

At very low frequencies, pulsar timing arrays (PTAs) are being used to search for the GW background from the merging population of massive black hole binaries (MBHBs). No detection has yet been made, but upper limits have been placed. Here we present results on what inference on the MBHB population can be learnt from present and possible future PTA results, and also compare current upper limits with astrophysical predictions, finding them to be fully consistent so far.

We also present a generic method for testing the consistency of a theory against experimental evidence in the situation where there is no strong viable alternative (for example GR). We apply this to BBH observations, finding them to be fully consistent with GR and also to Newton's constant of gravitation, where there is considerable inconsistency between measurements.

Acknowledgements

I have the pleasure of thanking a number of people. Most of all, my supervisor Alberto Vecchio for his great support, guidance and patience throughout and to Alberto Sesana my co-supervisor for many discussions and guidance. My collaborators as listed throughout this work, but in particular to Walter Del Pozzo for his mentorship. Also thank you to my examiners Rob Ferdman and Haixing Miao.

Thank you to all my friends in Birmingham who make our department a pleasure to be a part of, but especially to: Serena Vinciguerra as my desk neighbour and for our tea/coffee trips; David Stops for tea@1537, conkers, postcards, cryptic iconography, multiple monitor incidents... (and all things computer related); Anna Green, Sam Cooper and Andreas Freise, for all things outreach (and the most fun anyone can have when organising a museum exhibit!); Maggie 'the Martian and Queen of Outreach' Lieu for passing on her infectious enthusiasm for outreach; for pub trips to: Jim Barrett, Christopher Berry, Siyuan Chen, Chris Collins, Jack Gartlan, Kat Grover, Carl-Johan Haster, Matt Hunt, Max Jones, Sarah Mulroy, Coen Neijssel, Matt Robson, Daniel Töyrä, Alejandro Vigna-Gómez; and also to: Ben Bradnick, Daniel Brown, Simon Daley-Yates, Miguel Dovale, Will Farr, Sebastian Gaebel, Conner Gettings, Janna Goldstein, Ashley Jarvis, Aaron Jones, Ilya Mandel, Chiara Mingarelli, Alan McManus, James Nutt, Ian Stevens, Simon Stevenson, Laura Thomas, Gareth Thomas, Leon Trimble, John Veitch, Salvatore Vitale, Haoyu Wang; everyone at the Arboretum, but especially Les Wills for taking me under his wing; and to my family: my Uncle Fred, my brother Callum and to my parents Rob and Julie for their endless love and support.

This work was supported by the Science and Technology Facilities Council and the University of Birmingham. I am also grateful to the Royal Astronomical Society, the Institute of Physics and the Moreton University of Birmingham Travel Fund for funding assistance for attending conferences and workshops.

Contents

List of Figures

List of Tables

Statement of Originality

Chapter 1 is an introduction to the work presented here. It is written by me, but based entirely on review material which is referenced throughout the text.

Chapter 2 is reproduced in it's published form from Middleton et al. (2016). The work was done in collaboration with Walter Del Pozzo, Will Farr, Alberto Sesana and Alberto Vecchio. I led the code writing with assistance from co-authors, wrote the initial draft of the paper and edited later drafts

The main body of Chapter 3 is reproduced from Middleton et al. (2017) which is currently under consideration by Nature Communications. This was in collaboration with Siyuan Chen, Walter Del Pozzo, Alberto Sesana and Alberto Vecchio. For this work I have made contributions towards the code and analysis, wrote the initial draft and assisted editing for the paper text and produced some of the figures. Section 3.1 of chapter 3 summarises work which carried out in two papers (Chen et al., 2017a,b, of which I am also an author) and (Chen et al., 2017c) which are relevant to the chapter.

Chapter 4 contains work in progress from Del Pozzo et al. (in prep.), led by Walter Del Pozzo and with Alberto Vecchio, Ilya Mandel and Jonathan Gair. In this work I have contributed towards discussions, early versions of the code and ran the analysis for Newton's constant of gravitation.

In appendix A, I have included the transcript for Chen et al. (2017a) (and erratum: Chen et al. (2017b)), which was led by Siyuan Chen and in collaboration with Walter Del Pozzo, Alberto Sesana and Alberto Vecchio. For this work I have made contributions towards the code, some text and editing for the paper and made contributions for some of the figures.

Any errors throughout are entirely of my own doing.

Other work I have contributed towards

In this thesis I present work on astrophysical inference from pulsar timing array results and also on a generic means of testing the validity of a theory given multiple independent tests. As part of the LIGO-Virgo Scientific Collaboration, I have also contributed towards the following work throughout the course of my PhD.

Parameter Estimation for Binary Neutron-star Coalescences with Realistic Noise during the Advanced LIGO Era Berry et al. (2015)

Authorship: C. P. L. Berry, I. Mandel, H. Middleton, L. P. Singer, A. L. Urban, A. Vecchio, S. Vitale, K. Cannon, B. Farr, W. M. Farr, P. B. Graff, C. Hanna, C-J. Haster, S. Mohapatra, C. Pankow, L. R. Price, T. Sidery, J. Veitch

Contribution: Running analyses for this mock data challenge and assisted with collation of results.

Parameter Estimation on Gravitational Waves from Neutron-star Binaries with Spinning Components Farr et al. (2016)

Authorship: B. Farr, C. P. L. Berry, W. M. Farr, C-J. Haster, H. Middleton, K. Cannon, P. B. Graff, C. Hanna, I. Mandel, C. Pankow, L. R. Price, T. Sidery, L. P. Singer, A. L. Urban, A. Vecchio, J. Veitch, S. Vitale

Contribution: Running mock data challenge analyses.

Full author LIGO-Virgo Scientific Collaboration publications

Contribution: Running data analysis for gravitational wave signals/candidates as part of the Compact Binary Coalescence Parameter Estimation working group towards Abbott et al. (2016b), Abbott et al. (2016i), Abbott et al. (2016j), Abbott et al. (2016d), Abbott et al. (2016g) Abbott et al. (2017b)

Glossary

BBH Binary Black Hole

BHNS Black Hole Neutron Star (binary)

BNS Binary Neutron Star

CBC Compact Binary Coalescence

EMRI Extreme Mass Ratio Inspiral

EPTA European Pulsar Timing Array

FAST Five hundred metre Aperture Spherical Telescope

GR General Relativity

GW Gravitational Wave

GWYYMMDD Gravitatioanl Wave observed on date YYMMDD

GWSB Gravitational Wave Stochastic Background

H1 LIGO Hanford Observatory

IMBH Intermediate Mass Black Hole

IPTA International Pulsar Timing Array

ISCO Innermost Stable Circular Orbit

K-L Kullback-Leibler (divergence)

L1 LIGO Louisianna Observatory

LIGO Laser Interferometer Gravitational-wave Observatory

LISA Laser Interferometer Space Antenna

LVTYYMMDD LIGO-Virgo Trigger observed on date YYMMDD

MCMC Markov Chain Monte Carlo

MBH Massive Black Hole

MBHB Massive Black Hole Binary

NANOGrav North American Nanohertz Observatory for Gravitational Waves

PDF Posterior Density Function

PPTA Parkes Pulsar Timing Array

PTA Pular Timing Array

SKA Square Kilometre Array

TOA Time Of Arrival

Virgo gravitational wave observatory in Italy

Chapter 1

Introduction

This chapter provides a review and introduction to the material contained in later chapters. The text is written by me, but the contents is based entirely on the work of others who I have referenced throughout the chapter.

1.1 Setting the scene

Gravitational waves (GWs) are ripples in space-time predicted by Einstein over 100 years ago (Einstein, 1916, 1918) as a consequence of his theory of general relativity (GR). Produced by time-varying quadrupolar motion (Peters and Mathews, 1963), the observation of GWs provides a means to learn about the 'dark universe' complimentary to the electromagnetic spectrum. Although evidence for their existence was conclusive (Hulse and Taylor, 1975; Taylor and Weisberg, 1989; Weisberg and Taylor, 2005), they had not been detected until 2015, when the LIGO (Laser interferometer Gravitational-wave Observatory; Aasi et al., 2015) observed a signal from the inspiral and merger of two black holes (Abbott et al., 2016b). Since then, several more observations of binary black hole mergers (Abbott et al., 2016g, 2017b,e) have been made with

LIGO as well as with the Virgo gravitational wave observatory (Acernese et al., 2015). In the future, observations at multiple GW frequencies and from other source types will allow for the study of astronomical phenomena that may be inaccessible via other means.

In the following sections, we introduce the properties of gravitational waves (section 1.2) and describe the sources (section 1.3), with a particular emphasis on compact binary coalescences and the mergers of massive black hole binaries in sections 1.4 and 1.5. In section 1.6, we briefly cover the instruments and observations used to detect and search for gravitational waves with a review of observations and upper limits placed so far in section 1.7. Section 1.8 describes some of the techniques and methods used within this work. In chapter 2 we cover the astrophysical inferences possible on the merging population of massive black hole binaries using pulsar timing array searches for the stochastic gravitational wave background. Chapter 3 considers the current state-of-the-art pulsar timing array observations and the consequences of this for astrophysical predictions with the use of a more extended model than that used in chapter 2. Chapter 4 describes a well-known generic method for combining multiple experiments testing the same theory and applies this method to tests of general relativity and Newton's constant of gravitation.

1.2 Stretch and squash

Here we introduce some of the basic properties of gravitational waves (see, for example, Shapiro and Teukolsky, 1983; Flanagan and Hughes, 2005; Maggiore, 2008). Linearised gravity is an approximation to Einstein's equations in which the assumption of only

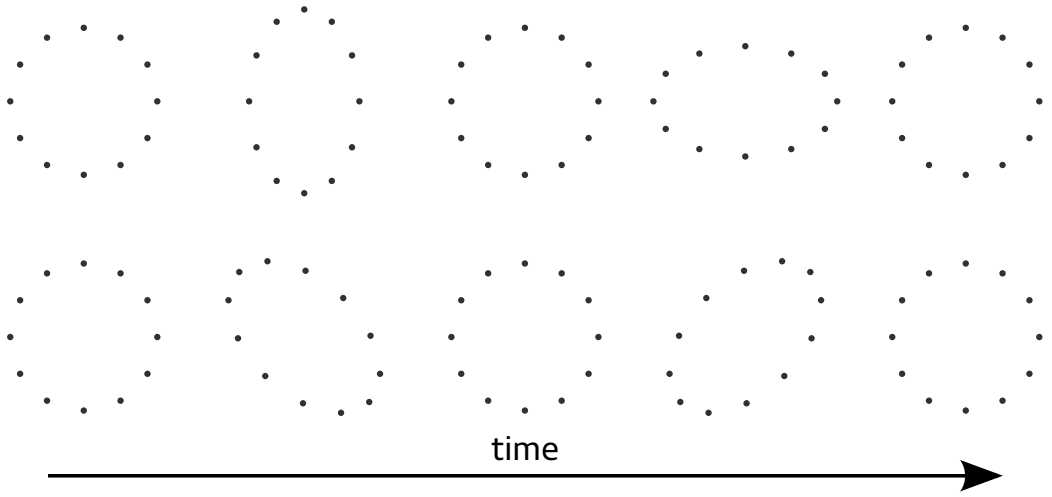


Figure 1.1: Showing the effect of a gravitational wave travelling into or out of the page on a set of test masses shown by the black dots. The test masses are origninaly aranged in a ring, and the change in their positions over time is show from left to right at fractions of the gravitational wave period (from left to right, fractions of 0, $1/4$, $1/2$, $3/4$, 0 of the period are shown). The top row shows the plus polarisation over time from left to right (h_+) and the bottom row shows the cross polarisation ($h_\times$).

small metric perturbations $h_{\mu\nu}$ from the flat space-time metric $\eta_{\mu\nu}$ is made,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \qquad |h_{\mu\nu}| \ll 1. \tag{1.1}$$

The tensor $h_{\mu\nu}$ describes time-varying oscillations to the space-time metric (ripples in the space-time curvature) which propagate at the speed of light, *i.e.* the gravitational waves. There are two independent GW polarisations: 'plus' h_+ and 'cross' $h_\times$. If propagating in the z direction, the 'stretch and squash' effect of a GW is in the x–y plane as illustrated by figure 1.1 for a single gravitational wave period in each polarisation. The assumption is also made that once the GWs have left their source vicinity, the radius of curvature of the background space-time they travel through is large in comparison to the gravitational wavelength λ_{gw}.

Two ways to think of the effect of GWs, and their measurement, are either by

looking for the change in separation between test masses (a distance measurement), or by measuring the change in frequency of a pulsed signal from some emitting source (a timing measurement). However the reality is that the effect is the same in both cases. In terms of distances, if ξ_j is the separation between two free test particles, a GW propagating in a direction perpendicular to the line of sight between them will produce a small relative acceleration. This results in a small change in their separation of

$$\delta\xi_j = \frac{1}{2} h_{jk}^{TT} \xi_k, \tag{1.2}$$

where h_{jk}^{TT} (using the transverse-traceless gauge) contains the two GW polarisations ($h_{xx}^{TT} = -h_{yy}^{TT} = h_+$ and $h_{xy}^{TT} = h_{yx}^{TT} = h_\times$). The gravitational wave strain amplitude h can be written as the change in separation over the separation itself,

$$\frac{\delta\xi}{\xi} \sim h. \tag{1.3}$$

Measurement of this effect is covered in more detail in section 1.6.1. For the latter case of a timing measurement, we imagine a pulsing or lighthouse-like signal with some period P_{pulse}. A change in the separation between source and receiver will change the light travel time and therefore the arrival time of the pulses, so that the observed frequency of pulsation ($\nu = 1/P_{\text{pulse}}$) is altered as

$$\frac{\delta\nu}{\nu} \sim h. \tag{1.4}$$

Further details on searches for gravitational waves via this means will be covered in section 1.6.2.

The strain amplitude of a GW signal is dependent on the source mass M in quadrupole motion (or how much mass-energy is converted into GWs) and decays as the inverse of

the distance r to the source. As an order of magnitude value,

$$h \sim \frac{M}{r}\left(\frac{v}{c}\right)^2 \quad \left(\sim \frac{R_{sch}}{r}\frac{v^2}{c^2}\right), \tag{1.5}$$

where v in the characteristic velocity of the source system and

$$R_{sch} = \frac{GM}{c^2}, \tag{1.6}$$

is the Schwarzschild raduis for a mass M (G and c are Newton's constant and the speed of light, respectively). From equation 1.5 we can see that any terrestrial source will have minuscule strain ($h \lll 1$). Therefore, as the creation of a GW source in the lab with sufficient amplitude for detection is nigh on impossible, we must turn to astronomical objects.

Here we provide some rough scales for h for different types of sources, which will be covered in more detail in the next sections. For an object like GW150914, the first observation of a binary black hole merger (Abbott et al., 2016b), the black holes each had a mass of ~ 30 $M_\odot$ and a gravitational wave frequency of ~ 150 Hz, which gives an orbital separation of ~ 350 km and an orbital velocity of $\sim 2/3c$. The source was at a distance of ~ 450 Mpc from the Earth, which using equation 1.5, gives a rough estimation of $h \sim 10^{-21}$. At the other end of the black hole mass scale, massive black holes (expected to form at the centres of merging galaxies) may have orbital frequencies of several years. For a massive black hole binary with component masses of $\sim 10^9$ $M_\odot$ on a ~ 10 yr orbit, their separation would be ~ 0.03 pc, corresponding to an orbital velocity of $\sim 0.06c$. If a distance of ~ 100 Mpc is chosen, then this would produce $h \sim 10^{-15}$.

1.3 Sources of gravitational waves

As we have seen in section 1.2, in order to produce GWs with a high enough amplitude for observation to be feasible, sources need to be of high mass and moving at relativistic speeds. In this section we briefly describe some of the astronomical sources of GWs (see also Cutler and Thorne, 2002, for an overview of sources). A particular focus on massive black hole binaries as GW sources relevant to this work will be covered in section 1.5.

1.3.1 Binary Inspiral / Merger

Compact Binary Coalescence

Compact binaries coalescences (CBCs) – involving binary black holes (BBHs), binary neutron stars (BNS) and black hole-neutron star binaries (BHNS) – provide sources with high masses and velocities as they inspiral and approach merger. Their signal increases in frequency and amplitude as the binary separation decreases, and this expected 'chirp-like' waveform for such a source can be approximated from GR (see figure 1.8 and for example Abbott et al., 2016c, for a comparison of waveforms used in the analysis of LIGO sources).

The first observational evidence of GWs came from radio observations of the Hulse-Taylor binary (Hulse and Taylor, 1975), a double neutron star binary where one of the objects appears to the Earth as a radio pulsar. The emission of GWs from an isolated binary such as this carries away energy, and therefore the separation of the components reduces. Observations for the Hulse-Taylor system carried out over many years (Weisberg and Taylor, 2005) show that the shrinkage of the orbital period matches extremely well to that expected if the system were emitting gravitational waves as predicted by GR to within $\sim 0.2\%$. Figure 1.2 shows the evolution of the system over 30 years of observations with the curved line labelled 'General Relativity Prediction'

showing how the system would be expected to evolve due to GW emission.

CBCs were considered one of the most promising sources (Abbott et al., 2016h) for LIGO's (section 1.6, Aasi et al., 2015) first GW detection. Indeed, the first observation of GWs passing through the Earth, GW150914, was made on 14 September 2015 from the inspiral and merger of two black holes, each with a mass of around 30 $M_{\odot}$ (Abbott et al., 2016b). Since this first observation, three other BBH signals, plus another candidate event, have been observed (Abbott et al., 2016g, 2017e). These observations are summarised in section 1.7.1.

Massive black hole binaries

Whilst LIGO has observed GWs from stellar mass BHs, massive black holes are also expected to produce GWs at much lower frequencies than those currently detectable by LIGO (and LISA in the future, see section 1.6). Massive black hole binaries as sources of gravitational waves will be covered in detail in section 1.5.

Other classes of inpirals – mergers

Although not discussed in this work, here we briefly mention other GW sources of this category. For example, the merger of a stellar mass object with an intermediate mass black hole (IMBH) of > 100 $M_{\odot}$ (see Mezcua, 2017, for a recent review and references therein). The observation of an inspiral like this would not only confirm the existence of IMBHs, but would also enable further tests of GR and the study of globular clusters, where they are expected to reside (Haster et al., 2016).

Another GW inspiral source are extreme mass ratio inspirals (EMRI) in which a small stellar mass object (of $\sim 0.5 - 50$ $M_{\odot}$) inspirals (likely with eccentric orbits, Barack and Cutler, 2004) into a massive black hole found at the centres of galaxies (Amaro-Seoane et al., 2007; Gair et al., 2013). The detection of these events will not only be of interest

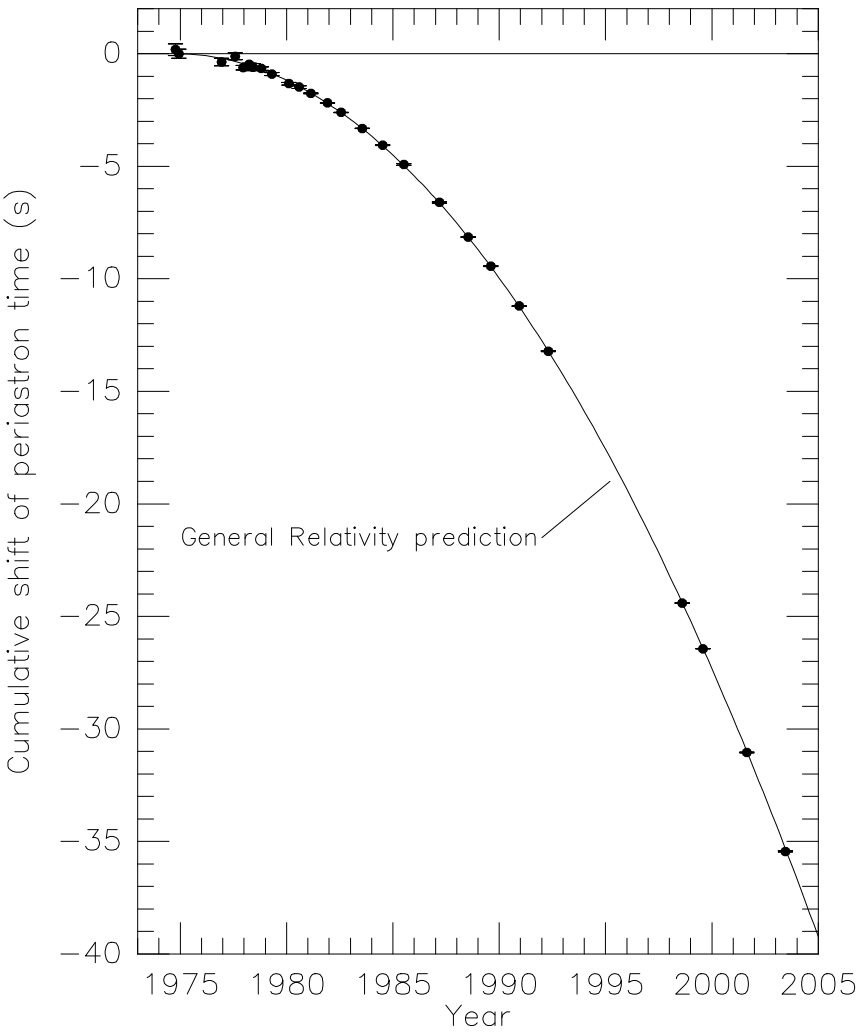


Figure 1.2: Showing the cumulative advance of periastron in the Hulse-Taylor binary over ~ 30 years of observations. As the orbit decays, the binary will return to its periastron position sooner on each orbit, leading to a cumulative shift in the periastron time. The points show the measurements of the system along with corresponding uncertainties. The curve shows the prediction for the evolution of the system if gravitational wave emission is taking place as described by general relativity The flat horizonal line at zero represents what would be expected if there were no orbital decay taking place. Reproduced from Weisberg and Taylor (2005).

astrophysically, but will provide another testing ground for GR, as these objects should make many orbits in the band of a space-based observatory like LISA (Amaro-Seoane et al., 2017; Babak et al., 2017, and see section 1.6).

1.3.2 Continuous Waves

Continuous wave refer to an on-going GW signal, such as a sine wave. An example of such a source would be a nearby (inside our galaxy) rotating neutron star with some amount of asymmetry, such as a mountain on its surface (Jones, 2002). No observation of a continuous source has been made so far, however upper limits have been placed on the GW strain amplitude in the range between 10^{-24} – 10^{-25} with LIGO (Abbott et al., 2017c, 2016e). Targeted searches of known objects have also been carried out by LIGO, with upper limits placed on the low-mass x-ray binary Scorpius X-1 (Abbott et al., 2017d).

1.3.3 Burst

Burst signals (Andersson et al., 2013), like CBCs, are transients. Whilst a CBC search in LIGO can use well modelled templates, searches for burst signals are 'unmodelled' to allow for a versatile range of possible signals. An example of such a source could be a supernovae explosion, however the versatility of the unmodelled search means it also is sensitive to unknown sources and CBCs – it was in fact a burst pipeline which first picked up the BBH merger GW150914 (Abbott et al., 2016a).

1.3.4 Stochastic background

A GW stochastic background is produced by many overlapping signals which cannot be individually resolved. Upper limits have been placed on the stochastic background from

both stellar-mass objects at frequencies accessible by LIGO (Abbott et al., 2016f) and also from massive black hole binaries (*e.g.*, Verbiest et al., 2016; Arzoumanian et al., 2016; Shannon et al., 2015; Lentati et al., 2015). The latter of these will be covered in more detail in later sections of this chapter, and inference from searches for the massive binary black hole population is the topic of chapters 2 and 3.

1.4 Some useful relations for gravitational waves and compact binaries

In this section we reproduce a few of the useful formulae and relations for GWs and circular binary inspirals (*e.g.* Maggiore, 2008; Shapiro and Teukolsky, 1983).

In the case where the GW wavelengths are much larger than the size of the source, the solution to the linearised Einstein equation for the GW strain h from an emitter at distance r is the quadrupole formula

$$h_{\mu\nu} = \frac{2}{r}\frac{G}{c^4}\ddot{Q}_{\mu\nu}, \tag{1.7}$$

where $Q_{\mu\nu}$ is the mass quadrupole moment, $Q_{\mu\nu} = \int \rho x_\mu x_\nu d^3x$ for a source with mass density $\rho(x_i, M)$. For a GW source of mass M and size R, the GW strain h scales as

$$h \sim \frac{G}{c^4}\frac{MR^2\omega^2}{r}, \tag{1.8}$$

where ω is the orbital frequency of the source. If considering a binary system of masses m_1, m_2 in a circular orbit, some substitution into equation 1.8 and including the correct

constant factors leads to

$$h = \left(\frac{32}{5}\right)^{1/2} \frac{(G\mathscr{M})^{5/3}}{r\,c^4}\,(\pi f_{gw})^{2/3}, \tag{1.9}$$

where $\mathscr{M}$ is the chirp mass

$$\mathscr{M} = \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}}, \tag{1.10}$$

and f_{gw} is the GW frequency (which is twice the orbital frequency f of the binary).

A number of useful relations can be computed using equation 1.9. The gravitational wave luminosity ($L = \mathrm{d}E_{gw,L}/\mathrm{d}t$) is

$$\frac{\mathrm{d}E_{gw,L}}{\mathrm{d}t} = 4\pi r^2\,F = 4\pi r^2\,\frac{\pi}{4}\,\frac{c^3}{G}\,f^2h^2, \tag{1.11}$$

$$= \frac{32}{5}\,\frac{\pi^{10/3}G^{7/3}}{c^5}\,(\mathscr{M} f_{gw})^{10/3}, \tag{1.12}$$

where F is the flux. Using Kepler's relation between the total mass of a system ($M = m_1 + m_2$), separation a, and period P of a binary

$$\frac{P^2}{a^3} = \frac{4\pi^2}{GM} \tag{1.13}$$

the change in orbital energy can be calculated as

$$\frac{\mathrm{d}E_{gw,orb}}{\mathrm{d}t} = -\left(\frac{G^2\mathscr{M}^5}{32}\right)^{1/3}\,\frac{2}{3}\,\omega_{gw}^{-1/3}\,\frac{\mathrm{d}\omega_{gw}}{\mathrm{d}t}, \tag{1.14}$$

where ω_{gw} is the angular GW frequency ($\omega_{gw} = 2\pi f_{gw}$). Therefore, simply by equating equations 1.12 and 1.14, an order of magnitude estimate of the change in GW frequency of a binary over time can be made,

$$\frac{\mathrm{d}f_{gw}}{\mathrm{d}t} = \frac{96}{5}\,\pi^{8/3}\,\left(\frac{G\mathscr{M}}{c^3}\right)^{5/3}\,f_{gw}^{11/3}. \tag{1.15}$$

Integrating of equation 1.15 leads to

$$t_c = \frac{5}{96} \left(\frac{c^3}{G\mathscr{M}} \right)^{5/3} (\pi f_{gw})^{-8/3}, \tag{1.16}$$

giving the time to merger for a binary system evolving under GW emission.

The final maximum frequency of a binary inspiral is decided by the separation that the two components can reach before rapid plunge and merger which takes place at the innermost stable circular orbit (ISCO). The ISCO radius R_{isco} for a non-spinning black hole of mass M_{bh} is given by

$$R_{isco} = \frac{6GM_{bh}}{c^2}, \tag{1.17}$$

therefore, again using equation 1.13, the ISCO frequency for an equal mass binary is given by,

$$f_{gw,isco} = \frac{1}{\pi\, 6\sqrt{6}} \frac{c^3}{GM_T}. \tag{1.18}$$

Some typical values of the ISCO frequency are: $f_{gw,isco} \sim 10^2$ Hz for a BBH of total mass ~ 60 M$_\odot$ (similar to GW150914); $f_{gw,isco} \sim 10$ Hz for an intermediate mass black hole binary with total mass ~ 200 M$_\odot$; and $f_{gw,isco} \sim 10^{-6}$ Hz for a massive black hole binary of total mass $\sim 10^9$ M$_\odot$.

1.5 Gravitational waves & massive black hole binaries

In this section we cover some of the relevant details on massive black holes (MBHs) as sources of gravitational waves that will be useful for later chapters, including a description of the population of merging black hole binaries (MBHBs) producing a stochastic background of gravitational wave radiation at low frequencies.

1.5.1 Massive black hole binaries

Most galaxies are believed to host MBHs at their centres with masses in the range of $\sim 10^6$ – 10^9 $M_\odot$ (see for example, Magorrian et al., 1998; Ferrarese and Merritt, 2000; Gebhardt et al., 2000; Kormendy and Ho, 2013). Evidence of this has been observed within our own galaxy from the trajectories of stars in the Galactic Centre around Sagittarius A*, a black hole estimated to have a mass of $\sim 4 \times 10^6$ $M_\odot$ (Gillessen et al., 2009; Ghez et al., 2008; Eisenhauer et al., 2005). In other galaxies evidence for the existence of central MBHs has been seen from quasar observations (with, for example the Sloan Digital Sky Survey; Pâris et al., 2017; Abolfathi et al., 2017).

Galaxies have also been observed to interact and merge with one another. For example, Centaurus A could be the product of a merger between an elliptical galaxy with a small galaxy (Tubbs, 1980). Hierarchical galaxy mergers throughout the course of cosmic time provide a convincing model for their growth and evolution (White and Rees, 1978). It is likely that mergers are frequently taking place throughout the history of the universe over the life of galaxies (Mundy et al., 2017; Lotz et al., 2011; Bell et al., 2006) – small galaxies grow by merging with other small galaxies at early times and these mergers continue through cosmic history as illustrated in figure 1.3. During these mergers, it is also reasonable to expect that the central black holes within each galaxy will form a binary and eventually merge, so that MBH growth goes hand-in-hand with the growth of the host galaxy (Kauffmann and Haehnelt, 2000; Croton et al., 2006).

During a galaxy merger, the two black holes will sink to the centre of the merger remnant via dynamical friction and interaction with the stars in the environment (Chandrasekhar, 1943a,b,a). The typical timescale for dynamical friction is on the scale of a few million years, given by (Begelman et al., 1980),

$$t_{df} \sim \frac{6 \times 10^6}{\log N} \left(\frac{v_c}{300\text{km s}^{-1}} \right) \left(\frac{r_c}{100\text{pc}} \right)^2 \left(\frac{m_2}{10^8\ M_\odot} \right)^{-1} \text{yr}, \tag{1.19}$$

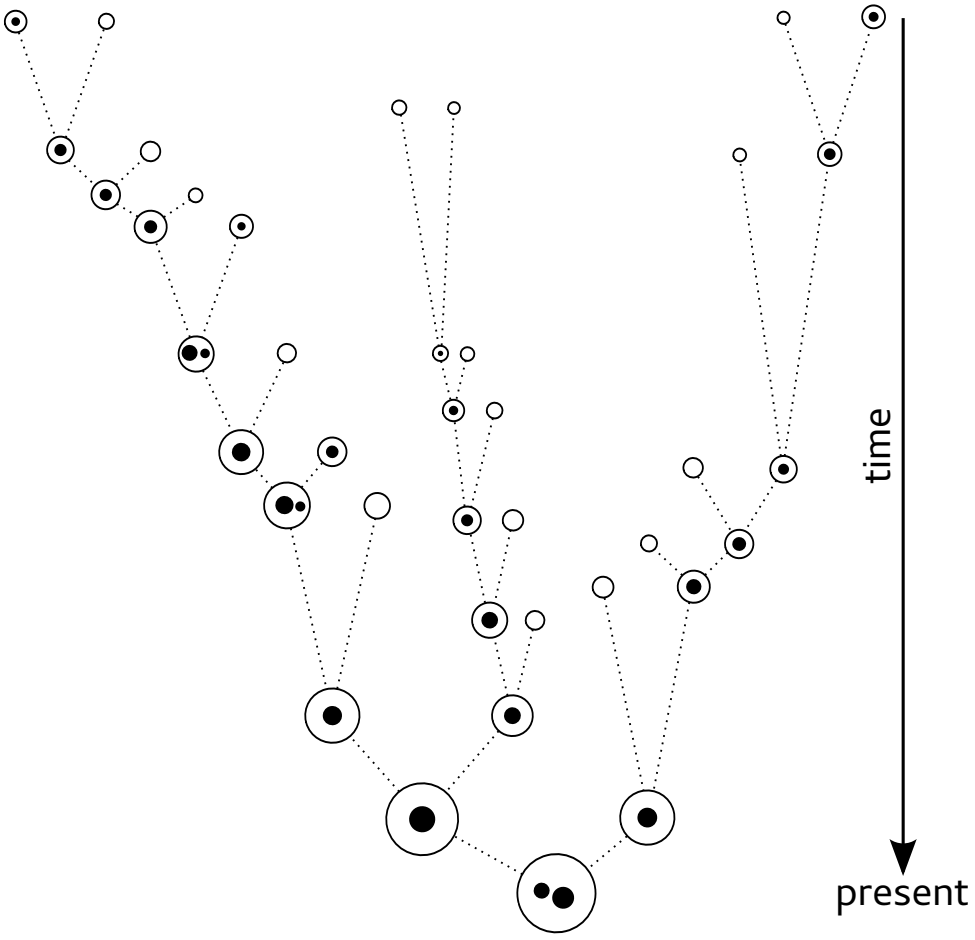


Figure 1.3: The galaxy merger tree, showing the growth of galaxies (circles) and their central black holes (black dots) via hierarchical mergers. At the top (earlier in time) small galaxies are merging with other small galaxies in the 'branches' and as time passes (moving down the diagram) mergers of larger objects take place reaching the 'trunk' at the present time. The galaxies increase in mass with each merger as do the central black holes. Image adapted from one by Marta Volonteri (University of Michigan).

where the central core containing N stars has a velocity dispersion v_c and radius r_c, and m_2 is the mass of the smaller black hole. As the black hole's separation decreases, further shrinking of the binary can be achieved via three-body scattering with individual stars (Quinlan, 1996; Mikkola and Valtonen, 1992). However, over time the stars in the vicinity of the black hole become depleted so that the number of stars within the binary's loss cone (*i.e.*, in orbital paths which take them into the radius of tidal disruption or capture of the binary – see figure 1.4) becomes fewer and fewer. As these stars are depleted, the binary's rate of shrinkage may slow and merger models can struggle to provide an explanation for the binary evolution after it reaches a separation of $\sim$ 1pc, known as the 'final parsec problem' (Milosavljević and Merritt, 2003). How to bridge this gap between the last parsec and the point at which radiation reaction becomes dominant is still uncertain, however several possible explanations have been discussed.

One means to continue the shrinkage of a MBHB is to provide some means of repopulating the loss cone so that stars can continue to interact with the binary. The efficiency of loss-cone repopulation can be improved if the galaxy is non-spherical or triaxial (Yu, 2002; Preto et al., 2011; Khan et al., 2011; Vasiliev et al., 2014, 2015; Sesana and Khan, 2015), and indeed a non-symmetric merger product from the interaction of two galaxies is very plausible. Another possibility is that if a subsequent galaxy merger takes place producing a triplet, the interaction with a third massive black hole component could speed up the binary merger (Hoffman and Loeb, 2007; Bonetti et al., 2016, 2017a,b). Gaseous disks may also provide a possible speed-up mechanism for the shrinkage of the binary (*e.g.* via interaction with a counter-rotating disk Schnittman and Krolik, 2015; Nixon et al., 2011; Cuadra et al., 2009), however this may not be enough in all cases, for example in the case of high mass, low redshift binaries (Dotti et al., 2015) or if the mass in the gas disk is too low (for example, Lodato et al., 2009, show that the mass in the disk needs to be at least comparable to the secondary, lighter black

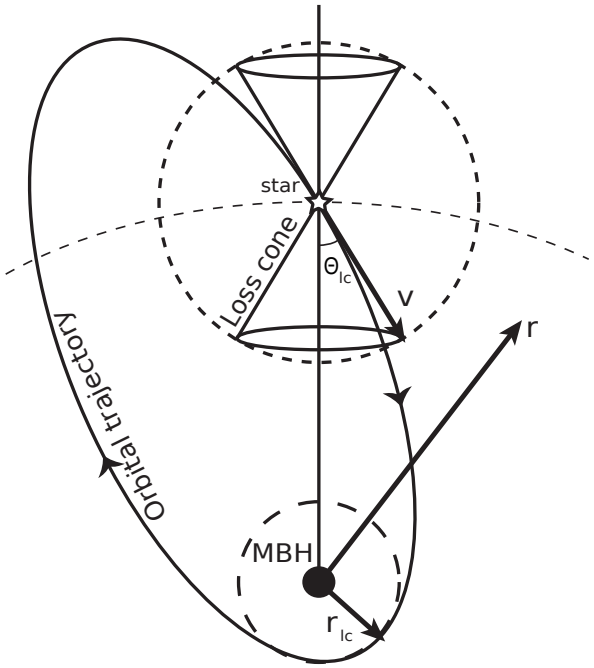


Figure 1.4: Representing the loss cone for a star orbiting around a massive black hole (labelled MBH) with velocity v. If the star's orbit lies inside the loss cone, it will bring it into the radius of tidal disruption or capture for the black hole r_{lc}. Reproduced from Merritt (2013).

hole). In N-body simulations of galaxies, rotation of the galaxy has been seen to speed up coalescence timescales by a factor of 3 – 30 for co- and counter-rotating galaxies respectively (Holley-Bockelmann and Khan, 2015).

Observational evidence for the existence of MBHBs residing at the centre of galaxies has been seen in radio galaxy 0402+379, where a massive black hole pair with a projected separation of only 7.3 pc has been observed (Rodriguez et al., 2006). Observations of quasars with periodicity in their light curves can also provide evidence for the existence of MBHBs, for example the quasar OJ287 is a candidate binary with a ~ 12 year periodicity (Valtonen et al., 2008). Two other recent examples of quasars identified to have periodicities in their light curves have periods of 1884 ± 88 days (PG 1302102;

Graham et al., 2015b) and of 542 ± 15 days (PSO J334.2028+01.4075; Liu et al., 2015). Another, 111 candidate objects have been identified (Graham et al., 2015a) in the Catalina Real-time Transient Survey (Djorgovski et al., 2011). However, other explanations (than the existence of a MBHB) could be the cause of these periodicities, such as a warped accretion disk, and even the conclusion that these light curve observations indeed show periodicities has been cast into some doubt (Vaughan et al., 2016).

1.5.2 Gravitational waves from massive black hole binary inspirals

A MBHB will decrease in separation via interaction with the environment until GW radiation reaction becomes the dominant process, eventually leading to the merger of the binary. Taking a typical MBH to be around $\sim 10^9$ $M_\odot$, from equation 1.18, the frequency at the last stable orbit is $\sim 10^{-6}$ Hz. However, a typical value for the frequency at transition to GW emission is $\sim$ nHz (Chen et al., 2017c), so the inspiral phase can take $\sim$ Gyrs. As we expect there to be an isotropically distributed population of these sources throughout the universe, these long inspirals will produce a stochastic background of gravitational waves from MBHBs (see for example Rajagopal and Romani, 1995; Wyithe and Loeb, 2003; Jaffe and Backer, 2003; Sesana et al., 2008). By searching for the stochastic gravitational wave background, we can hope to learn about the underlying population of merging MBHBs and gain insight into the formation and evolution of galaxies (Sesana, 2013a; Burke-Spolaor, 2015).

1.5.3 Population of merging massive black hole binaries

The gravitational wave stochastic background (GWSB) from a population of merging MBHBs can be described as the sum of contributions from all sources, weighted by the source property distributions (*e.g.*, in mass and redshift). Here we reproduce the result

from Phinney (2001) to describe such a population.

As a starting point, we have the number of black hole merger remnants N per co-moving volume V_c, per unit time in the source frame t_r and logarithmic chirp mass interval $\log_{10}\mathscr{M}$. If n is the number of merger remnants per unit co-moving volume ($n = \mathrm{d}N/\mathrm{d}V_c$) then

$$N(z, \log_{10}\mathscr{M}) = \frac{\mathrm{d}^2 n}{\mathrm{d}t_r \, \mathrm{d}\log_{10}\mathscr{M}} = \frac{\mathrm{d}^2 n}{\mathrm{d}z \, \mathrm{d}\log_{10}\mathscr{M}} \, \frac{\mathrm{d}z}{\mathrm{d}t_r}, \tag{1.20}$$

where (Hogg, 1999)

$$\frac{\mathrm{d}z}{\mathrm{d}t_r} = \frac{1}{H_o(1+z)E(z)}, \tag{1.21}$$

$$E(z) = (\Omega_M(1+z)^3 + \Omega_\Lambda)^{1/2}, \tag{1.22}$$

and we assume standard cosmological values of $H_o = 70$ km s^{-1} Mpc^{-1}, $\Omega_M = 0.3$ and $\Omega_\Lambda = 0.7$ for Hubble's constant, mass density and dark energy density parameters, respectively.

If we define Ω_{gw},

$$\Omega_{gw}(f) = \frac{1}{\rho_c c^2} \frac{\mathrm{d}\rho_{gw}}{\mathrm{d}\ln f}, \tag{1.23}$$

as the present-day energy density in gravitational waves ρ_{gw} per logarithmic GW frequency interval (which from here we label f, dropping the gw subscript) divided by the critical mass energy density ($\rho_c c^2$), then the total present day energy density in gravitational waves is the integration over frequency,

$$\varepsilon_{gw} \equiv \int_0^\infty \rho_c c^2 \Omega_{gw}(f) \frac{\mathrm{d}f}{f} \equiv \int_0^\infty \frac{\pi}{4} \frac{c^2}{G} f^2 h_c^2(f) \frac{\mathrm{d}f}{f}, \tag{1.24}$$

where $\rho_c = 3H_o^2/8\pi G$ is the universe critical density, and $h_c(f)$ is the characteristic

amplitude of the gravitational wave spectrum over a logarithmic frequency interval. In an isotropic and homogeneous universe, ε_{gw} must also be equal to the sum of contributions from all sources, dividing by $(1+z)$ to account for redshifting. Here we integrate over both redshift and logarithmic chirp mass due to our choice of population model in equation 1.20 (however the choice could be made to use *e.g.* the mass ratio of the binary instead),

$$\varepsilon_{gw} = \int_0^\infty \int_0^\infty \int_0^\infty N(z, \log_{10}\mathscr{M}) \frac{1}{1+z} \frac{\mathrm{d}E_r}{\mathrm{d}f_r} f_r \frac{\mathrm{d}f}{f} \, \mathrm{d}z \, \mathrm{d}\log_{10}\mathscr{M}, \tag{1.25}$$

where $N(z, \log_{10}\mathscr{M})$ is the number density of mergers at redshift z and with chirp mass $\mathscr{M}$ (whose form will be covered later), and $(\mathrm{d}E_r/\mathrm{d}f_r)f_r$ is the energy emitted in gravitational waves in the frequency range f_r to $f_r + \mathrm{d}f_r$ (where the subscript r refers to the source rest frame). Therefore, by equating these two equations (1.24 and 1.25), we can compute the characteristic amplitude's dependence on the cosmic population of MBHBs as

$$h_c^2(f) = \frac{4G}{\pi c^2} \frac{1}{f^2} \int_0^\infty \int_0^\infty N(z, \log_{10}\mathscr{M}) \frac{1}{1+z} \frac{\mathrm{d}E_{gw}}{\mathrm{d}f_r} f_r \, \mathrm{d}z \, \mathrm{d}\log_{10}\mathscr{M}. \tag{1.26}$$

Division of equations 1.12 and 1.15 gives us an expression for energy in GW per frequency bin for a single binary as

$$\left.\frac{\mathrm{d}E}{\mathrm{d}f_r}\right|_{f_r = f(1+z)} = \frac{(\pi G)^{2/3}}{3} \frac{\mathscr{M}^{5/3}}{f_r^{1/3}}, \tag{1.27}$$

and by substitution into equation 1.26

$$h_c^2(f) = \frac{4G^{5/3}}{3\pi^{1/3}c^2} \, f^{-4/3} \int_0^\infty \int_0^\infty N(z, \log_{10}\mathscr{M}) \frac{\mathscr{M}^{5/3}}{(1+z)^{1/3}} \, \mathrm{d}z \, \mathrm{d}\log_{10}\mathscr{M}, \tag{1.28}$$

we can describe the expected gravitational wave background from a population of sources, $N(z, \log_{10}\mathscr{M})$.

The function $N(z, \log_{10}\mathscr{M})$ can be chosen to be representative of the chirp mass and redshift distribution of the massive black hole population, as well as the overall merger rate density. This will be covered in chapter 2. The relation shown by equation 1.28 ($h_c \sim f^{-2/3}$) describes the GWSB from a population of MBHBs evolving via radiation reaction only in circular orbits. If binaries are in eccentric orbits, the spectrum will no longer follow this simple $f^{-2/3}$ power law. The effect of, and consequences for, their detection will be covered in more detail within chapter 3.

1.6 Instruments

In this section we provide a short discussion on instruments and techniques used to observe and search for gravitational waves. The frequency of a GW depends on the source and, like the electromagnetic spectrum, different detectors and techniques are needed to access different frequency regimes of the GW spectrum. Figure 1.5 shows diagrammatically the spectrum of sources found at different frequencies and the instruments used to search for and detect them. These sources will be covered in more detail throughout this section.

We focus on ground-based interferometer detectors, such as Advanced LIGO and Advanced Virgo (Aasi et al., 2015; Acernese et al., 2015) and pulsar timing arrays (*e.g.*, Verbiest et al., 2016), covering the high and very low frequency GW sources respectively. Here we do not cover future space-based interferometric observatories such as LISA (Laser Interferometer Space Antenna; Amaro-Seoane et al., 2017), which will be able to bridge this frequency gap (see figure 1.5). LISA is currently planned for launch during the 2030s and is proposed to have a triangular configuration of three identical

satellites in formation, linked by 2.5×10^6 km laser arms (Amaro-Seoane et al., 2017). A technology demonstration mission, LISA Pathfinder, was launched in 2015 with the aim of testing the technology necessary for a space-based observatory and has been very successful (Armano et al., 2016).

1.6.1 Ground-based observatories

Ground-based observatories are laser interferometers which measure the relative change in separation between test masses caused by passing gravitational waves. The first observation of gravitational waves from the merger of a binary black hole system was made in 2015 by the twin LIGO detectors (Aasi et al., 2015; Abbott et al., 2016b). The two observatories are located in Livingston (L1), Louisiana and Hanford (H1), Washington (see the map inset in figure 1.6).

A simplified diagram of the instrument is shown in figure 1.6. In the interferometer, laser light is split into two arms by a beam splitter. The light in each of the arms (which are 4 km long in LIGO) enters a Fabry-Perot cavity where the mirrors, acting as the test masses, are suspended from fibres to reduce environmental noise. The two light paths recombine at the beam splitter and the measured beam at the photo-diode is the result of the intereference of these two beams. Any motion of the test mass mirrors which causes a relative change in the distance travelled by the light in each arm will cause a change in relative phase at the beam splitter and in turn change the intensity at the photo-diode.

At the time of the first detection, the observatories H1 and L1 were operating in their advanced configuration during 'Observing Run 1' (September 2015 – January 2016). Inset (b) of figure 1.6 shows the sensitivity of the instrument around the time of the first detection. One measure of the astrophysical sensitivity of a detector is the 'horizon distance' – the distance at which a face-on, overhead BNS source (with typical 1.4 $M_\odot$ neutron star components) would be detectable at a threshold signal-to-noise

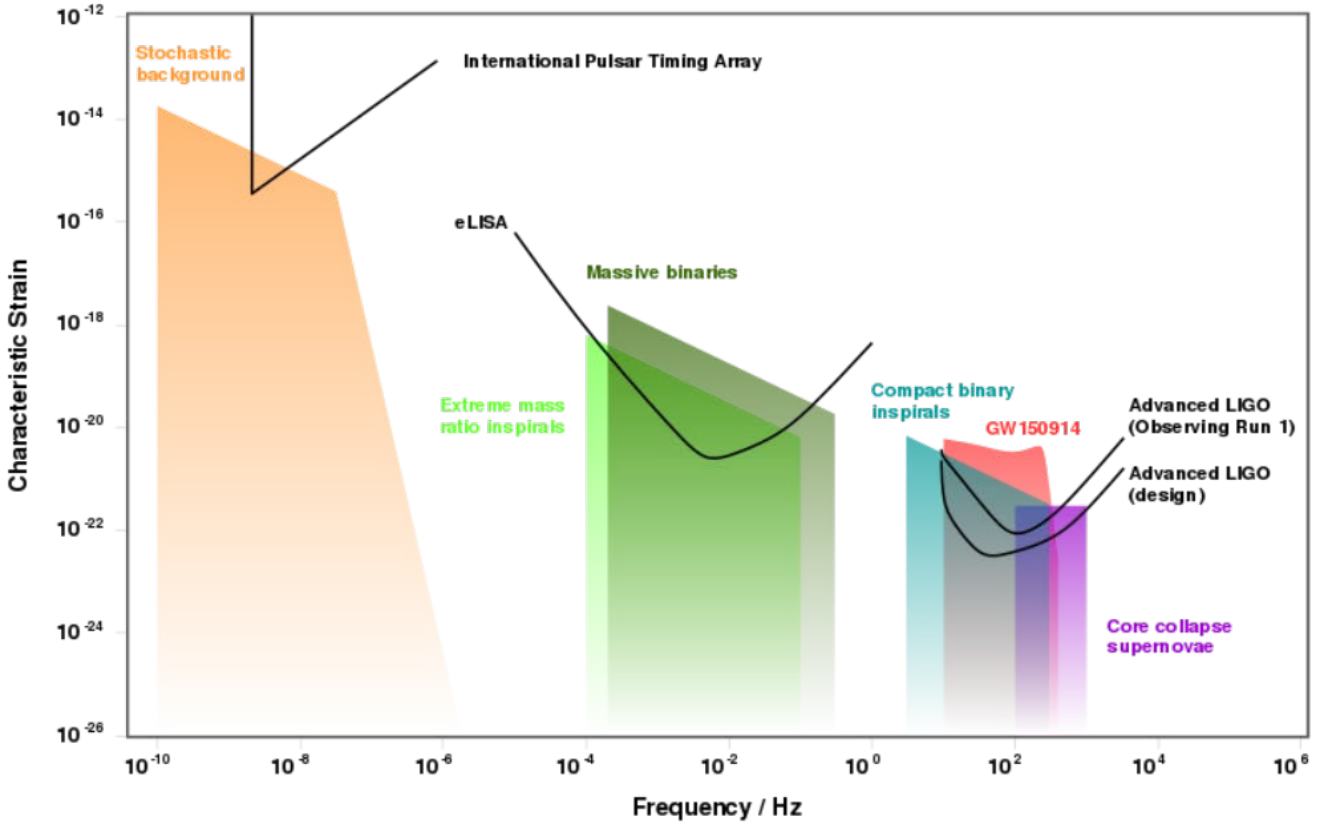


Figure 1.5: The gravitational wave spectrum. Gravitational wave sources are shown as coloured blocks and the instruments are represented by black lines indicating their sensitivity curves. On the far left, the low frequency sources are from the stochastic background due to massive black hole mergers (orange) accessible via pulsar timing arrays (*e.g.*, Verbiest et al., 2016). The IPTA sensitivity curve is based on 20 pulsars, observed for 15 years with a cadence of 2 weeks and a timing precision of 1×10^{-7}s (for details see, Moore et al., 2015b,a) In the centre are massive black hole binaries (dark green) and extreme mass ratio inspirals (light green) accessible via a space-based observatory such as LISA (Amaro-Seoane et al., 2017). At high frequency individual transient sources are shown including compact binary inspirals (blue), such as GW150914 (red) (Abbott et al., 2016b) and supernovae events (purple) accessible by ground-based interferometers like Advanced LIGO (Aasi et al., 2015) and Advanced Virgo (Acernese et al., 2015). Here, LIGO sensitivities shown for both its advanced configuration during the first observing run 'Observing Run 1' (September 2015 – January 2016), and also at design. Plotted using *GW Plotter*. See `rhcole.com/apps/GWplotter/` for plotting and for further details of how this plot is constructed, see Moore et al. (2015a).

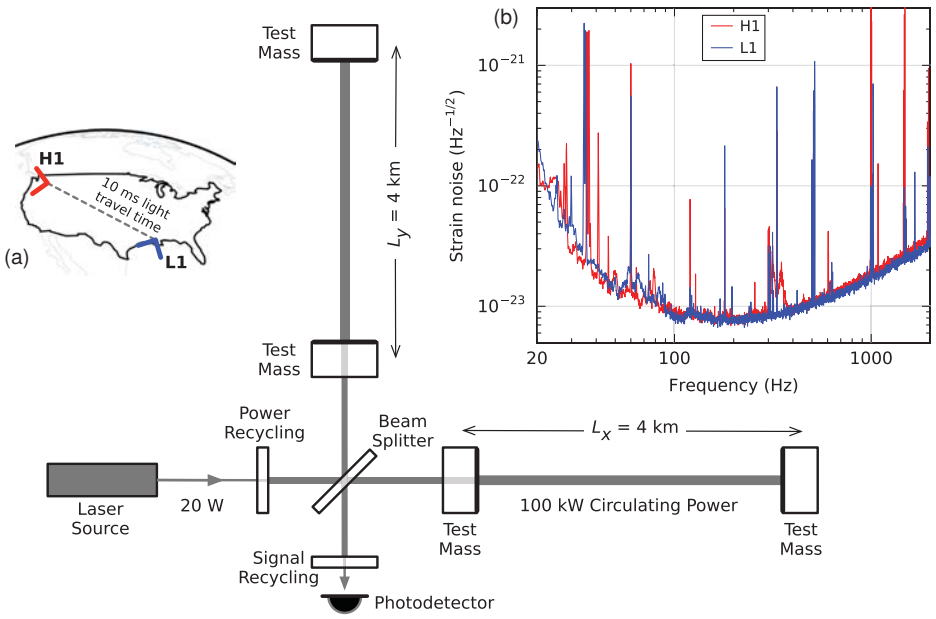


Figure 1.6: A simplified schematic of a ground-based gravitational wave observatory like Advanced LIGO. Inset (a): the location and orientation of the twin LIGO detectors in Hanford, Washington (H1) and Livingston, Louisiana (L1). Inset (b): the instrument noise over frequency for H1 and L1 near the time of the first observation of GWs (GW150914) in terms of the equivalent GW strain amplitude. The thin spike features in the spectrum are due to calibration lines, vibrational modes of the mirror suspension fibres and the electric power grid (60 Hz) harmonics. Reproduced from Abbott et al. (2016b).

ratio of 8 (Finn and Chernoff, 1993; Abbott et al., 2016h). As this is constructed for the optimal source configuration, another measure called the BNS range gives the distance to which these signals can be observed once averaged over source position and orientation. During Observing Run 1, the detectors were operating at a BNS range of 60–80 Mpc and this was improved to 60–100 Mpc during Observing Run 2 (30 November 2016 – 25 August 2017; Abbott et al., 2016h). The 3 km observatory, Advanced Virgo (located in Italy Acernese et al., 2015) recently joined the network on 1 August 2017, and the first triple-detector observation of a binary black hole merger was made on 14 August 2017 (GW170814; Abbott et al., 2017e), leading to a marked improvement on source localisation constraints (see right panel of figure 1.9). There is also a smaller 600 m detector, GEO 600 (Affeldt et al., 2014), located in Germany.

Looking to the future, these instruments will be joined by a broader network of GW observatories across the globe. KAGRA, an underground cryogenic-cooled interferometer is under construction in the Kamioka mine, Japan (Aso et al., 2013) and LIGO-India is being planned (Iyer et al., 2011). Farther ahead, next generation ground-based observatories are being conceived, including the 40 km Cosmic Explorer (Abbott et al., 2017a) and the Einstein Telescope (Sathyaprakash et al., 2012), a triangular interferometer with a broader range of frequency sensitivity. These additions and upgrades to the GW network will not only provide improved sensitivities, but will also improve the constraints placed on source parameters (Fairhurst, 2011; Gaebel and Veitch, 2017).

1.6.2 Pulsar Timing Arrays

Moving to the very low frequency, as we described in section 1.5, it is expected that there is a $\sim$ nHz gravitational wave stochastic background from the merging population of massive black hole binaries (with masses in the range $\sim 10^6$ – 10^9 $M_\odot$). Radio pulsars can be used to search for these GWs (Sazhin, 1978; Detweiler, 1979).

Pulsars are neutron stars whose rotation axis is offset from the magnetic axis, from which they emit a beam of radiation at radio wavelengths. For those who lie in the path of the radio beam, as the star rotates, a lighthouse-like beacon is observed. The first pulsar was discovered in 1967 with a rotation period of 1.337 s (Hewish et al., 1969), but it was the discovery of a pulsar with a $\sim$ millisecond rotation period (Backer et al., 1982) and others like it, which has led to the use of pulsars as extremely accurate clocks. Pulsar systems have also provided a testing ground for GR, with systems such as the Hulse-Taylor binary (Hulse and Taylor, 1975, and see section 1.3.1) and the double pulsar, where both components appear as radio pulsars (Burgay et al., 2003; Kramer et al., 2006).

Following a recent review by Hobbs and Dai (2017) (and references therein), for a pulsar located at a distance D from the Earth, the effect of a GW with strain h_{ij} on the observed pulse frequency ν is given by

$$\frac{\delta\nu}{\nu} = -H^{ij}\left[h_{ij}\left(t_e, x_e^i\right) - h_{ij}\left(t_e - D/c, x_p^i\right)\right], \tag{1.29}$$

where $\delta\nu$ is the change in the frequency, the geometrical term H^{ij} depends on the relative positions of the GW source, the Earth and the pulsar, and $x_{\{e,p\}}$ describes the position of the Earth and the pulsar at the times $t_{\{e,p\}}$ when the GW passes each of them. In reality it is the time of arrival (TOA) of the pulses which is measured. After observing a pulsar over a time period of many years, the TOA of each pulse can be predicted with high accuracy. The residuals $R(t)$, at time t from the start of observations, are the difference between the predicted and observed TOA,

$$R(t) = -\int_0^t \frac{\delta\nu}{\nu}\mathrm{d}t. \tag{1.30}$$

It is this measurement which contains the imprint of the GWSB. An approximate

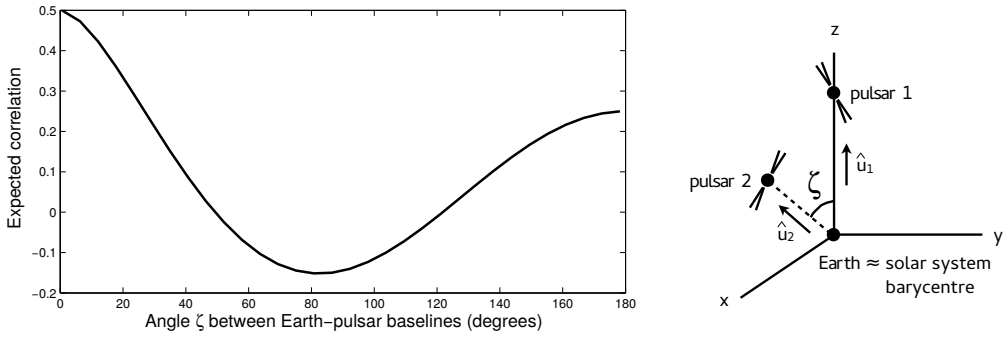


Figure 1.7: The Hellings and Downs curve (Hellings and Downs, 1983). Left: the expected correlation between pulsar pairs at different angular separations ζ due to an isotropic gravitational wave stochastic background. Right: the geometry of pulsar angular separations. Images reproduced from Jenet and Romano (2015).

value for induced timing residual from an individual source binary of total mass M, at luminosity distance d_L, radiating at a GW frequency of f would be (Jenet et al., 2009),

$$R(t) \sim 10\text{ns} \left(\frac{1\text{Gpc}}{d_L}\right) \left(\frac{M}{10^9\ \text{M}_\odot}\right)^{5/3} \left(\frac{10^{-7}\ \text{Hz}}{f}\right)^{1/3}, \tag{1.31}$$

however the overall signal is the GWSB from many MBHB sources.

The GWSB signal will be common to all pulsars. One pulsar is not enough to confirm a detection; a pulsar timing array (PTA) of pulsars across the sky is needed. The effect of the GWSB on a PTA is to create correlations in the residuals which are dependent on the angular separation of pulsar pairs on the sky. The predicted correlation for an isotropic GWSB is the Hellings and Downs curve (Hellings and Downs, 1983; Jenet and Romano, 2015) which is shown in figure 1.7, however if the background is anisotropic, the expected correlation will differ (Mingarelli et al., 2013).

It is however not enough to monitor just any set of pulsars, the chosen pulsars themselves must be 'good timers'. Most pulsars have rotation periods of $P \sim 0.5$ s, and their periods increase over time at a rate of $\dot{P} \sim 10^{-15}$ s/s (Lorimer, 2008). However, millisec-

ond pulsars are much more stable clocks with $P \leq 30$ ms and $\dot{P} \leq 10^{-19}$ s/s (Lorimer, 2008), providing suitable timing precision for the construction of a pulsar timing array (Detweiler, 1979). Therefore the common or garden pulsar is not sufficient for gravitational wave searches, and we must turn to millisecond pulsars for their timing precision.

There are, however, several other phenomena aside from gravitational waves which can cause a delay in the measured arrival times, leading to variations in the timing residuals, and make the GWSB signal extremely difficult to isolate (Lentati et al., 2015). Several of these effects can be specific to individual pulsars, and therefore should not introduce additional correlations between pulsars (as is expected from the GWSB), however it is important to be able to understand these effects to find the GWSB. Examples include: astrometric properties; pulsars in binaries (*e.g.* Doppler shifts and relativistic effects); spin-down of pulsars; pulse profile variability Brook et al. (2014); Brook (2015); Shannon et al. (2014); pulse jitter (Liu et al., 2012); the effect of changes in density of the interstellar medium along the line of sight on pulse dispersion (Lee et al., 2014; Keith et al., 2013); pulsar glitches (*e.g.*, McKee et al., 2016; Cognard and Backer, 2005) and red noise (*e.g.* Perrodin et al., 2013). There are also effects which will be common to all pulsars such as: the timing standards used in the analysis; the choice of solar system ephemeris (Hobbs and Dai, 2017); and local radio frequency interference, largely from terrestrial activity (*e.g.* An et al., 2017). For techniques used in pulsar timing analysis, see for example, Hobbs et al. (2006); PINT (2017); Taylor et al. (2017).

Currently there are several observational efforts using radio telescopes around the globe to time pulsars across the sky in search of an inkling of the GWSB. They are the European Pulsar Timing Array (EPTA: Lentati et al., 2015), the Parkes Pulsar Timing Array (PPTA: Shannon et al., 2015) and the North American Nanohertz Observatory for

Gravitational Waves (NANOGrav: Arzoumanian et al., 2016). Together they form the International Pulsar Timing Array consortium (IPTA: Verbiest et al., 2016). Observations for a given pulsar are typically made with a cadence of $\sim 1/\text{week}$ and many pulsars have now been observed for over a decade, so the sensitivity band for PTAs is in the range $\sim 10^{-6}$–10^{-9} Hz (Hobbs and Dai, 2017), where it is expected that the gravitational wave stochastic background from MBHBs will be found.

In the future, several new observatories will be added to the IPTA effort, including MeerKAT (Foley et al., 2016) in South Africa, FAST (Five Hundred Metre Aperture Spherical Telescope; Nan et al., 2011) and QiTai (Xu et al., 2017), both in China; as well as the Square Kilometre Array (SKA; Dewdney et al., 2009), split between South Africa and Australia. These observatories will enable the discovery of more millisecond pulsars as well as high precision timing (Janssen et al., 2015).

1.7 Observations and upper limits for binary black holes

Here we provide details of observations of gravitational waves from binary black holes made so far and the current best upper limits where detections have not yet been made. We only detail here the sources relevant to this work, which are stellar mass binary black hole observations by LIGO (Aasi et al., 2015) and upper limits on the massive binary black hole population searched for by pulsar timing arrays (Verbiest et al., 2016).

1.7.1 Detections

Advanced LIGO and Advanced Virgo have observed gravitational waves passing through the Earth from the mergers of binary black holes (BBHs). The sources GW150914, GW151226, GW170104 and GW170814 (Abbott et al., 2016b,d, 2017b,e) were all observed with low false alarm rates (based on how often a noise excursion in the detector

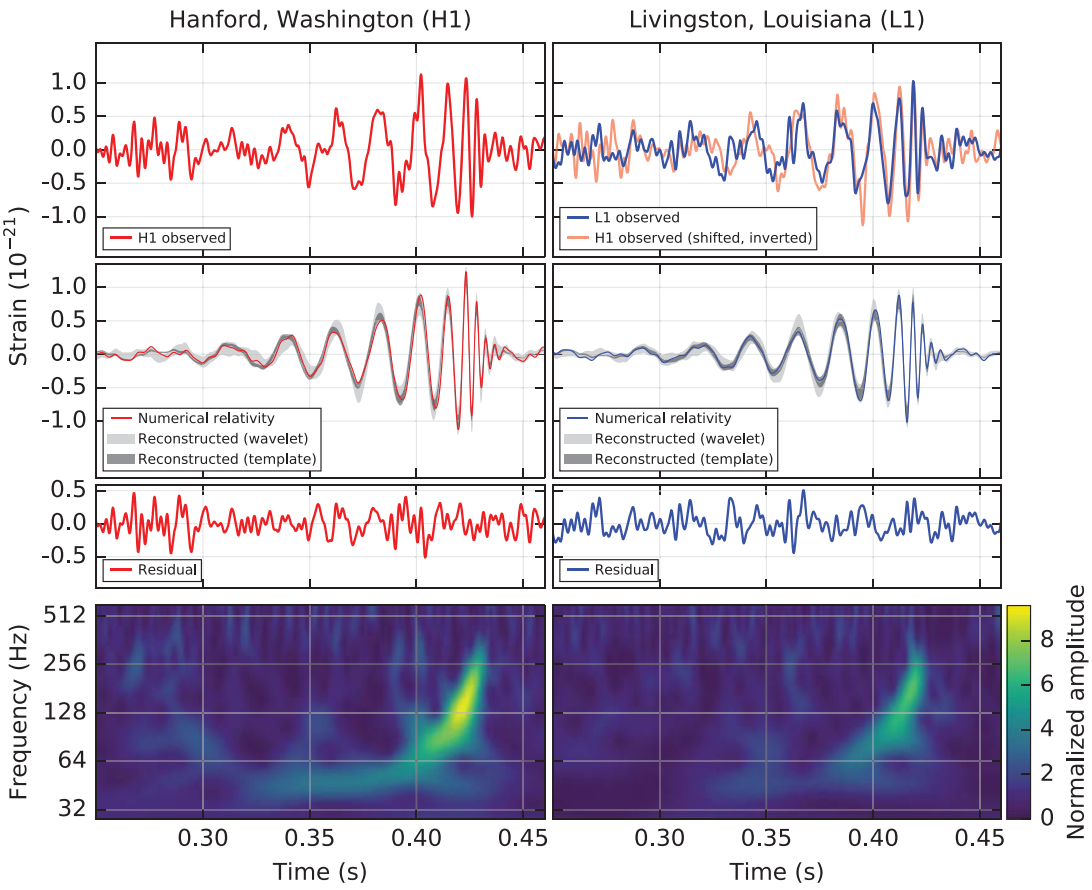


Figure 1.8: The first observation of gravitational waves by Advanced LIGO (GW150914), detected on 14 September 2015. The left and right panels show the event in the twin detectors LIGO Hanford (H1) and LIGO Livingston (L1), respectively. All times are shown relative to September 14, 2015 09:50:45 UTC. First row: the GW strain in H1 and L1 (with the H1 signal time-shifted and inverted for visual comparison). Second row: the 90% credible regions for two independent reconstructions of the waveform are shown using binary black hole template waveforms (dark grey) and sine-Gaussian wavelets (light grey) along with a numerical relativity waveform consistent with the recovered parameters for GW150914 (solid red lines) Third row: the residuals after the subtraction of the numerical relativity waveform. Fourth row: time-frequency representation of the strain data, where the signal can be clearly seen increasing in frequency over time. Reproduced from Abbott et al. (2016b).

might create a signal of the same strength) as shown in table 1.1. There is also a fifth BBH merger, LVT151012 which is described as a 'candidate event' due to it's higher false alarm rate (Abbott et al., 2016g). The observation of GW150914, the first BBH, is shown in figure 1.8, and the source properties for all events so far are summarised in table 1.1.

Figure 1.9 shows some selected results from the observations so far. The left panel shows the posterior distribution for the component masses of all but the latest BBH mergers. The right panel shows sky localisation. The sky area that the observations can be constrained to is important for electromagnetic follow-up observations. As only the twin LIGO detectors observed the first four BBH signals, their sky locations are poorly constrained, however the improvement upon the addition of Advanced Virgo to the network is very apparent in the comparative size of the localisation area for GW170814.

1.7.2 Upper limits on the massive black hole binary population

Generally speaking, before a technique is sensitive enough to make a detection, information can still be gained from placing upper limits based on the sensitivity of the instrument or experiment. This is the case for pulsar timing array searches for the gravitational wave stochastic background due to the merging population of massive binary black holes. Typically these are quoted as 95% upper limits on the characteristic strain at a frequency of 1 yr^{-1}, which we label $h_{\mathrm{yr},95\%}$. If the assumption of circular binaries driven by radiation reaction is taken, then the spectrum is a simple power-law characterisable by a single number, $h_c \propto f^{-2/3}$ (see also equation 1.28).

To date, the best PTA upper limit is held by the PPTA at $h_{\mathrm{yr},95\%} = 1 \times 10^{-15}$ (Shannon et al., 2015), with similar upper limits from the NANOGrav and EPTA at 1.5×10^{-15} and 3×10^{-15}, respectively (Arzoumanian et al., 2016; Lentati et al., 2015). These upper limits are shown by the dashed lines in figure 1.10. Recent unpublished work has

Event	GW150914	GW151226	LVT151012	GW170104	GW170814
Primary mass ($M_\odot$)	$36.2^{+5.2}_{-3.8}$	$14.2^{+8.3}_{-3.7}$	23^{+18}_{-6}	$31.2^{+8.4}_{-6.0}$	$30.5^{+5.7}_{-3.0}$
Secondary mass ($M_\odot$)	$29.1^{+3.7}_{-4.4}$	$7.5^{+2.3}_{-2.3}$	13^{+4}_{-5}	$19.4^{+5.3}_{-5.9}$	$25.3^{+2.8}_{-4.2}$
Chirp mass ($M_\odot$)	$28.1^{+1.8}_{-1.5}$	$8.9^{+0.3}_{-0.3}$	$15.1^{+1.4}_{-1.1}$	$21.1^{+2.4}_{-2.7}$	$24.1^{+1.4}_{-1.1}$
Total mass ($M_\odot$)	$65.3^{+4.1}_{-3.4}$	$21.8^{+5.9}_{-1.7}$	37^{+13}_{-4}	$50.7^{+5.9}_{-5.0}$	$55.9^{+3.4}_{-2.7}$
Final mass ($M_\odot$)	$62.3^{+3.7}_{-3.1}$	$20.8^{+6.1}_{-1.7}$	35^{+14}_{-4}	$48.7^{+5.7}_{-4.6}$	$53.2^{+3.2}_{-2.5}$
Effective inspiral spin parameter	$-0.06^{+0.14}_{-0.14}$	$0.21^{+0.20}_{-0.10}$	$0.0^{+0.3}_{-0.2}$	$-0.12^{+0.21}_{-0.30}$	$0.06^{+0.12}_{-0.12}$
Final black hole spin	$0.68^{+0.05}_{-0.06}$	$0.74^{+0.06}_{-0.06}$	$0.66^{+0.09}_{-0.10}$	$0.64^{+0.09}_{-0.20}$	$0.70^{+0.07}_{-0.05}$
Luminosity distance (Mpc)	420^{+150}_{-180}	440^{+180}_{-190}	1000^{+500}_{-500}	880^{+450}_{-390}	540^{+130}_{-210}
Source redshift	$0.09^{+0.03}_{-0.04}$	$0.09^{+0.03}_{-0.04}$	$0.20^{+0.09}_{-0.09}$	$0.18^{+0.08}_{-0.07}$	$0.11^{+0.03}_{-0.04}$
False alarm rate (yr^{-1})	$< 6 \times 10^{-6}$	$< 6 \times 10^{-6}$	0.37	$< 2 \times 10^{-5}$	$< 4 \times 10^{-5}$

Table 1.1: Source properties for observations of gravitational waves from merging binary black holes; GW150914, GW151226, LVT151012, GW170104 and GW170814. Reproduced from Abbott et al. (2016g), Abbott et al. (2017b) and Abbott et al. (2017e)

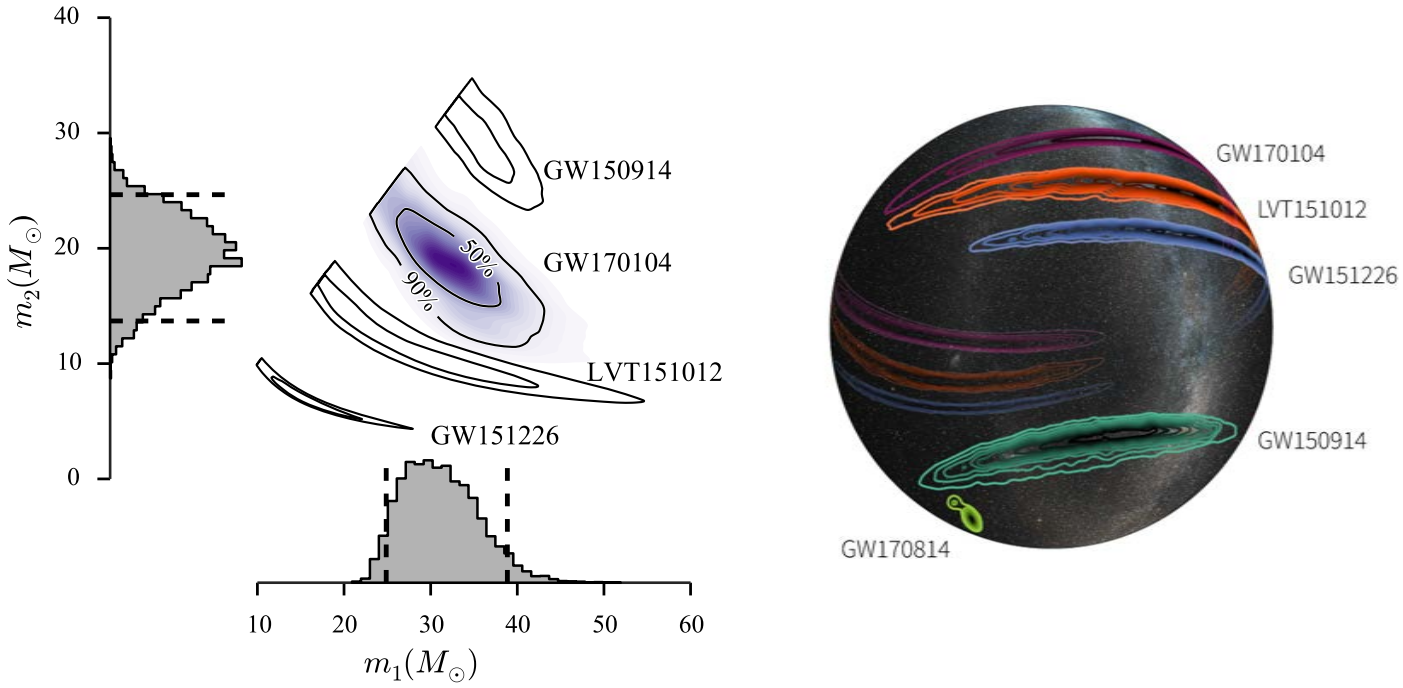


Figure 1.9: Some properties of the GW sources observed by LIGO and Virgo to date. Left: the two-dimensional posterior distribution for the source-frame component masses m_1, m_2 of the black holes for GW150914, LVT151012, GW1512126 and GW170104 (highlighted in blue), where the centre contours mark the 50% confidence band and the outer the 90% band. The contours follow the lines of roughly constant chirp mass (the well-measured mass parameter) and the diagonal hard cut-off is due to the choice of $m_1 > m_2$. The histograms show the marginalised one-dimensional posterior distributions for GW170104 only. Reproduced from Abbott et al. (2016g). Right: Constraints placed on the locations of the BBH signals observed by LIGO and Virgo to date. The improvement with the addition of the Advanced Virgo detector is very apparent as the most recent detection of GW170814 (lime green) shows. As more detectors are added to the network, the localisation of sources will improve. Image produced by LIGO/Caltech/MIT/Leo Singer/Milky Way image: Axel Mellinger.

however suggested that these results may be affected by the choice of solar system ephemeris and this is an active area of study.

Upper limits, although not as informative as detections, can still provide some information on the underlying population of merging massive black holes, and are now starting to reach the level of some theoretical predictions, as shown by figure 1.10. Our work on this is covered in detail within chapters 2 and 3.

1.8 Techniques

Throughout this work, we use a number of statistical analysis techniques. In this section we briefly describe some of the background for those that are used in later chapters (see also: Gregory, 2005; Jaynes, 2003; Mackay, 2003).

1.8.1 Bayesian Analysis

As a starting point we write down Bayes' theorem which at its heart, originates from the product rule,

$$P(A,B|C) = P(A|C)P(B|A,C), \tag{1.32}$$

$$= P(B|C)P(A|B,C). \tag{1.33}$$

Here, anything to the right of the bar | is assumed to be true and any two occurrences listed together (*e.g.* A,B) are joint probabilities: in the first line, $P(A,B|C)$ is the joint probability of A and B, conditional on C, $P(A|C)$ is the probability of A given C and $P(B|A,C)$ is the probability of B given A *and* C. The terms that make up the second line are similarly defined. By equating the expressions 1.32 and 1.33, we can write Bayes

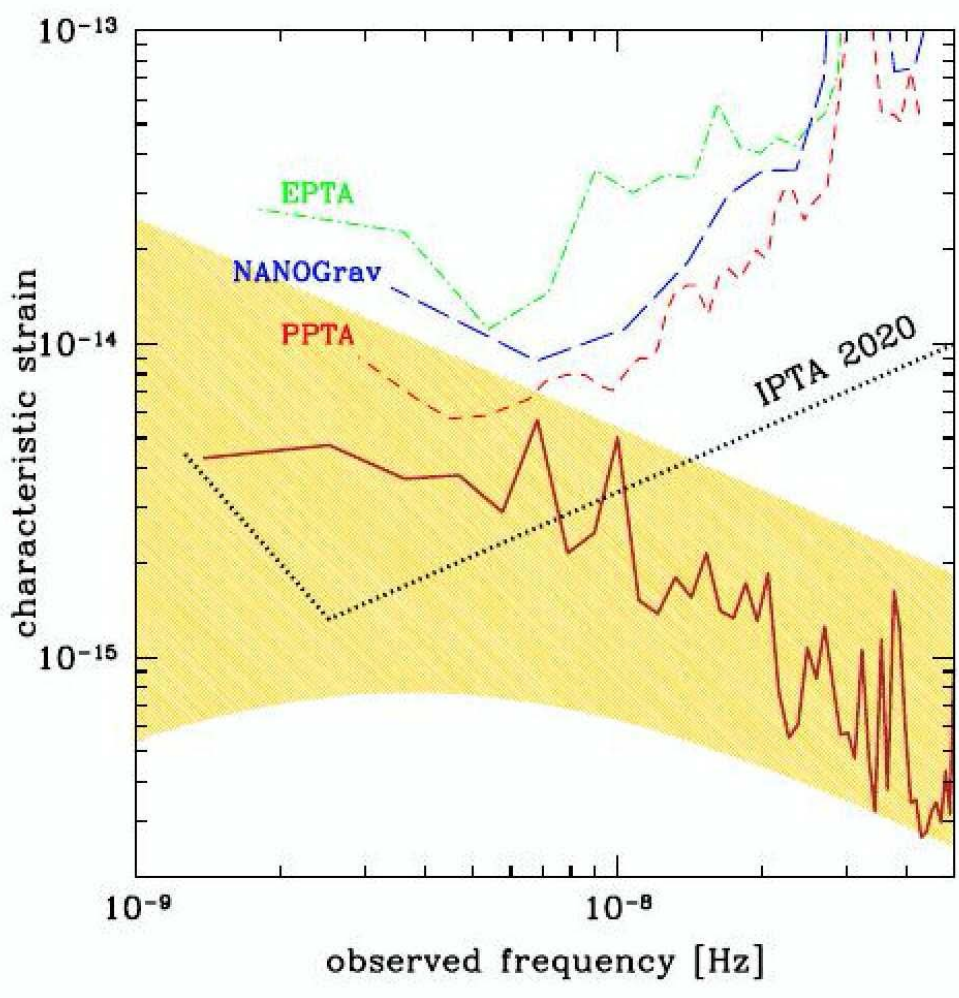


Figure 1.10: The characteristic strain amplitude for the pulsar timing array frequency window. The dashed lines show the current upper limits from the EPTA (green, dash-dot), NANOGrav (blue, long-dash) and PPTA (red, short-dash). A possible sensitivity which may be achievable by the IPTA by 2020 is shown by the dotted black line. A theoretical prediction from Sesana et al. (2016) for the GWSB is shown by the yellow-orange shaded band with just one possible realisation of the population shown by the jagged solid line. The width of the band is dependent on the uncertainty in the theoretical prediction. We see that the current upper limits are starting to probe the theoretical region. Image credit: A. Sesana, reproduced from Hobbs and Dai (2017).

theorem in its simplest form as

$$P(A|B,C) = \frac{P(A|C)P(B|A,C)}{P(B|C)}. \tag{1.34}$$

When doing Bayesian inference, our motivations usually fall into one of two classes: parameter estimation or model selection.

Parameter Estimation

In the case of parameter estimation, we assume the truth of a single model M, which is dependent on a number of parameters θ. Our objective is to find the posterior probability density function $P(\theta|d,M,I)$ for these parameters, given some data d and any other information we have I. So we rewrite Bayes' theorem as

$$P(\theta|d,M,I) = \frac{P(\theta|M,I)P(d|\theta,M,I)}{P(d|M,I)}. \tag{1.35}$$

Here $P(\theta|M,I)$ represents our prior knowledge on θ, $P(d|\theta,M,I)$ is the likelihood of the data given some combination of the parameters and assuming the truth of model M, and $P(d|M,I)$ is the evidence, which acts as a normalisation factor.

The likelihood is the probability of this dataset assuming the truth of the model and any prior information on the parameters. For N data points, the likelihood can be written as,

$$\mathcal{L} = P(d|\theta,M,I) = \prod_{i=1}^{N} P(d_i|\theta,M,I), \tag{1.36}$$

which for a Gaussian likelihood with mean μ and standard deviation σ would be

$$P(d|\theta,M,I) = \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(d_i-\mu)^2}{2\sigma^2}\right), \tag{1.37}$$

where we have also made the assumption that all the d_i have the same error σ.

It is often the case that we are only interested in the relative posterior of positions in parameter space, rather than the value itself. Therefore, the computation of the log-likelihood is sufficient and in this example would be

$$\ln \mathcal{L} \propto \sum_{i=1}^{N} \frac{(d_i - \mu)^2}{2\sigma^2}. \tag{1.38}$$

Once we have a likelihood, the next step is to specify the prior.

The prior is dependent on what we expect for the parameters and can take any form we choose. For example, an uninformative prior might be a uniform distribution giving equal weighting to all values: however, in reality the computation requires a parameter range to be specified. If for example, the parameter could cover several orders of magnitude, a prior might be chosen which is flat in the log to allow for even weighting of high and low values.

In parameter estimation, results are often represented as marginalised distributions, where we have integrated (or marginalised) over the other parameters of the model. An example of this is shown in the left panel of figure 1.9, where the two-dimensional distributions show the posterior marginalised over all other parameters except from the two component masses m_1 and m_2, whilst the one-dimensional histograms show the posterior marginalised over all but one parameter. Results for individual parameters are often quoted as credible bands (or credible intervals). For example, a 90% credible band would mean that 90% of the posterior lies within this band, *i.e.* $\int_b p(\theta|d,M,I) = 90\%$ where b is the integration region. This can be chosen as the central band, or equally as an upper or lower band, depending on the situation.

In practice, posteriors on parameters are often computed via sampling techniques, particularly for high dimensional parameter spaces. Some examples of widely used algorithms are described in section 1.8.2.

Model Selection

In the alternative case of model selection, there are a number of models M_j which we would like to weigh up against each other, again in light of data d and any information I that we already have. The model may or may not have parameters θ. Here we write Bayes' theorem as

$$P(M_j|d,I) = \frac{P(M_j|I)P(d|M_j,I)}{P(d|I)}, \tag{1.39}$$

where $P(M_j|I)$ is our prior belief in model M_j given some information I, $P(d|M_j,I)$ is the likelihood of getting data d with this particular model and $P(d|I)$ is the evidence, which we will call $\mathcal{Z}_i$ (see also equation 1.42). For each model, we would like to know the posterior for that model, given the data and any prior information – independently of the model parameters (if any), over which we marginalise

$$p(M_j|d,I) = \int_a^b \mathrm{d}\theta p(M_j,\theta|d,I), \tag{1.40}$$

where in this case, we have chosen a uniform prior bounded by a and b (specifying the region of integration).

For model selection, we would like to compare models A and B by computing the odds ratio,

$$\mathcal{O}_{A,B} = \frac{\mathcal{Z}_A}{\mathcal{Z}_B}\frac{P(M_A)}{P(M_B)} = \mathcal{B}_{A,B}\frac{P(M_A)}{P(M_B)}, \tag{1.41}$$

where $\mathcal{B}_{A,B}$ is the Bayes factor (the ratio of model evidences) between models A and B. In the case where we have equal prior belief in all models the odds ratio is equal to the Bayes factor. The odds ratio allows for the direct comparison of two models; however, it cannot provide a conclusive statement about the overall belief in a single model without comparison to another.

1.8.2 Sampling

When performing parameter estimation, we would like to be able to find the most favourable regions of parameter space by finding the highest values of the posterior distribution. In the case of a one or two parameter model, it might be convenient to simply create an evenly spaced grid of values in the parameter space and compute the posterior at each grid point. However, with a larger number of model parameters, this becomes extremely computationally expensive. It is more efficient to be able to explore the parameter space using stochastic sampling. Here, samples taken are weighted by the distribution itself – high probability areas are better sampled than low probability areas. The following sections describe two sampling methods which will be used in later chapters of this work.

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) in one of its simplest form is the Metropolis-Hastings algorithm (Metropolis et al., 1953; Hastings, 1970; Gregory, 2005). The basic structure of this technique is as follows:

1. Choose a starting position in parameter space θ_{i-1} and calculate the probability of this position $P(\theta_{i-1})$ based on a chosen likelihood and prior as described within section 1.8.1.

2. Select a trial position θ_{trial}, whose position depends only on the previous position θ_{i-1}, and calculate $P(\theta_{\text{trial}})$. This position could be, for example, drawn from a normal distribution with σ jump size, $\mathcal{N}(\theta_{i-1}, \sigma^2)$.

3. Decide whether to accept or reject θ_{trial} based upon comparison of the relative probability of the two positions and a random number, r which is uniformly drawn between 0 and 1 ($r = \mathcal{U}(0,1)$),

$$\text{if:} \quad \left(r < \frac{P(\theta_{\text{trial}})}{P(\theta_{i-1})}\right) \rightarrow \quad \text{accept trial:} \quad \theta_i = \theta_{\text{trial}}; \quad P(\theta_i) = P(\theta_{\text{trial}})$$

$$\text{else:} \quad \rightarrow \quad \text{reject trial:} \quad \theta_i = \theta_{i-1}; \quad P(\theta_i) = P(\theta_{i-1})$$

4. Repeat steps 2 and 3 to achieve the required number of samples.

The result of this algorithm is that a more probable position will always be accepted as the inequality, $P(\theta_{\text{trial}})/P(\theta_i) > 1$ is always true. However, in the case where a less probable position is drawn, $P(\theta_{\text{trial}}) < P(\theta_i)$ then it is possible to move to a lower probability position, enabling the exploration of the whole parameter space if the algorithm continues for a sufficiently long time. The result of this process is a joint posterior distribution for each of the parameters.

There are numerous ways to improve the efficiency of this method and to increase the rate of independent samples. For example, certain choices can be removed from the user, such as the choice of starting position in step 1 and the size of the jump to the trial position in step 2. An example of such a sampler used in later chapters of this work is `emcee` (Foreman-Mackey et al., 2013). This is an ensemble sampler, where a set of 'walkers' explore the parameter space together and jumps are decided based on the distribution of walkers, which becomes tighter and tighter as they explore the parameter space, finding the most likely areas.

Nested Sampling

Nested sampling is a means to compute multidimensional integrals. The method differs from that of MCMC in that it is designed to compute the evidence with the posterior distribution for the parameters being a by-product. The evidence $\mathcal{Z}$ is the integral of the

product of the prior and likelihood,

$$\mathcal{Z} = \int_{\Theta} p(\theta|M,I)p(d|\theta,M,I)\mathrm{d}\theta\,, \tag{1.42}$$

$$\approx \sum_{i=1}^{N} \mathcal{L}_i w_i\,, \tag{1.43}$$

where $\mathcal{L}_i$ is the likelihood $\mathcal{L}_i = p(d|\theta,M,I)$ and w_i is the 'weight' $w_i = p(\theta|M,I)d\theta$ from the prior.

In nested sampling, we start by randomly placing a set of N 'live points' distributed throughout the prior. Each point has some likelihood value $\mathcal{L}_j$ and lies on a contour X_j of equal likelihood value as shown by figure 1.11. The live points can be ordered by the amount of prior volume enclosed by the contour it lies on – the lowest likelihood contour encloses the largest volume and the highest likelihood contour encloses the smallest volume. Using this picture, the evidence can be expressed as a one-dimensional integral,

$$\mathcal{Z} = \int \mathcal{L}(X)\mathrm{d}X\,, \tag{1.44}$$

$$\approx \sum_{j} \mathcal{L}(X_j)\Delta X_j\,, \tag{1.45}$$

where ΔX_j is the distance between two contours $X_j - X_{j-1}$.

On each iteration, the live point lying on the lowest probability contour (X_1 in figure 1.11) is removed and replaced with a new point sampled from the remaining prior volume within that contour so that $X_{new}(L > L_1)$. Over time, the volume shrinks, building up detail of the integral at higher likelihoods. An example of such a sampler, which is used later in this work, is `cpnest` (Veitch et al., 2017).

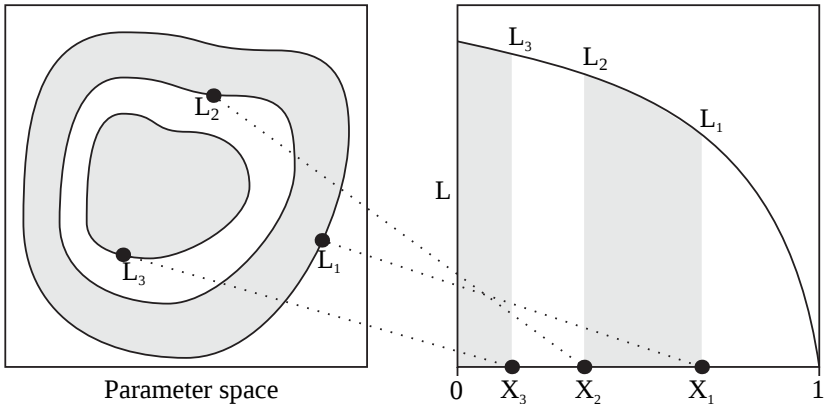


Figure 1.11: On the left, the parameter space is represented as a two-dimensional space in which three live points are placed, with likelihood values $L_{\{1,2,3\}}$. The contours $X_{\{1,2,3\}}$ on which each point lies are of equal likelihood. On the right is the one-dimensional representation of the evidence integral which is built up as the lowest likelihood live points are removed and replaced. Reproduced from Veitch and Vecchio (2010).

1.8.3 K-L Divergence

It can be very useful in a Bayesian analysis to have some indication of the amount 'we have learnt' after the addition of some data. The Kullback-Leibler (K-L) divergence, also called the relative entropy (Kullback, 1959; Mackay, 2003), provides a quantitative measure of the difference between two distributions over the same parameter x. For example if comparing the distribution $P(x)$ to another $Q(x)$, the K-L divergence is given by,

$$D_{\mathrm{kl}}(P||Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}, \tag{1.46}$$

which satisfies the inequality

$$D_{\mathrm{kl}}(P||Q) \geq 0. \tag{1.47}$$

The K-L divergence is zero only if the two distributions are identical ($P = Q$). In practice, when calculating the K-L divergence, if either distribution is zero ($P(x) = 0$ or

$Q(x) = 0$), then to avoid infinities we set $D_{\rm kl} = 0$

The result of this is a measure of how one probability distribution differs from a second probability distribution. We will use to compare our prior and posteriors in chapter 3, as a measure of how much our data has updated our prior.

1.9 Summary of chapters

In the following chapter 2, we report the results from Middleton et al. (2016) concerning what inference can be made on the massive black hole population with pulsar timing array upper limits and future detections. This work makes the assumption that the binaries follow circular orbits driven by radiation reaction alone, so that the $h_c \propto f^{-2/3}$ spectrum of equation 1.28 holds true.

In chapter 3, we provide a summary of work done by Chen et al. (2017c,a,b) in which a model is constructed to allow for eccentricity in the binaries. This model is then applied to the most stringent upper limit from the Parkes Pulsar Timing Array (Shannon et al., 2015) to discover what constraints may be placed on astrophysical predictions for the gravitational-wave background. This chapter is taken from Middleton et al. (2017).

In chapter 4 we present a work in progress for a generic test of theories whose underlying basis is the century-old Fishers combined probability test (Fisher, 1970). We apply this method to parametrised tests with gravitational-wave binary black hole observations and also to measurements of the Newtonian constant of gravitation.

Our conclusions are summarised in chapter 5.

Chapter 2

Astrophysical constraints on massive black hole binary evolution from Pulsar Timing Arrays

The work within this chapter is presented in the form of a paper published in Monthly Notices of the Royal Astronomical Society (Middleton et al., 2016).
The text is taken directly from the paper, with some minimal reformatting changes and removal of some repetition of the introductory material found in chapter 1. This work was done in collaboration with the co-authors (listed below). I led the code writing with assistance from co-authors, wrote the initial draft of the paper and edited later drafts.

Authorship:

Hannah Middleton, Walter Del Pozzo, Will M. Farr, Alberto Sesana, Alberto Vecchio.

Abstract

We consider the information that can be derived about massive black-hole binary (MBHB) populations and their formation history solely from current and possible future pulsar timing array (PTA) results. We use models of the stochastic gravitational-wave background from circular MBHBs with chirp mass in the range $10^6 - 10^{11}$ $M_\odot$ evolving solely due to radiation reaction. Our parametrised models for the black hole merger history make only weak assumptions about the properties of the black holes merging over cosmic time. We show that current PTA results place an upper limit on the black hole merger density which does not depend on the choice of a particular merger history model, however they provide no information about the redshift or mass distribution. We show that even in the case of a detection resulting from a factor of 10 increase in amplitude sensitivity, PTAs will only put weak constraints on the source merger density as a function of mass, and will not provide any additional information on the redshift distribution. Without additional assumptions or information from other observations, a detection *cannot* meaningfully bound the massive black hole merger rate above zero for any particular mass.

2.1 Introduction

Massive black holes (MBHs) reside at the centres of most galaxies (see *e.g.* Kormendy and Ho, 2013, and references therein), and are believed to have a central role in their evolution (see *e.g.* Volonteri, 2012, and references therein for a recent review). Mapping the population of MBHs, studying their properties, demographics, and their connection to the broader formation of structure is one of the open problems of modern astrophysics. This is however difficult to tackle, due to the large range of scales and the wide variety of physical processes involved. The MBH evolutionary path remains a highly debated

subject with many competing hypotheses still in play. Currently favoured hierarchical structure formation scenarios imply frequent galaxy mergers (White and Rees, 1978). As a result MBH binaries (MBHBs) should be quite common in the universe (Begelman et al., 1980; Volonteri et al., 2003). To-date there is no confirmed observed MBHB, although a number of candidates exist (see *e.g.* Dotti et al., 2012, and references therin) and tantalising claims have been recently made (Graham et al., 2015b; Liu et al., 2015; Graham et al., 2015a).

A means to survey MBHBs is through observation of the gravitational waves (GWs) that these systems generate as they inspiral towards their final merger. The accurate timing of an array of highly stable millisecond pulsars – a Pulsar Timing Array (PTA; Foster and Backer, 1990) – provides a direct observational means to probe the cosmic population of MBHBs on orbital timescales of order of several years. Astrophysical modelling suggests that the radiation emitted by an ensemble of MBHBs produces a GW stochastic background (GWSB) in the frequency range $\sim 10^{-9} - 10^{-7}$ Hz, where PTAs operate (Sesana et al., 2008, 2009; Ravi et al., 2012; Sesana, 2013b). Such a background affects the time of arrival of radio pulses in a characteristic fashion (Sazhin, 1978; Detweiler, 1979; Hellings and Downs, 1983), which can be used to discriminate the signal from a plethora of other undesired effects (Lentati et al., 2015).

Over the last decade, pulsar timing has placed progressively tighter constraints on gravitational radiation in this frequency regime (see, *e.g.* Jenet et al., 2006). More recently the three international consortia consisting of the PPTA, NANOGrav and the EPTA, which in collaboration form the IPTA (see section 1.6.2 and Shannon et al., 2015, 2013; Demorest et al., 2013; Arzoumanian et al., 2016; Lentati et al., 2015; van Haasteren et al., 2012, 2011; Hobbs et al., 2010) have used data from observations of unprecedented sensitivity to place constraints that are starting to probe astrophysically interesting regions of the parameter space (Sesana, 2013b).

In this chapter, we consider a GWSB produced by MBHBs in circular orbits losing energy and angular momentum purely through GW emission. We use an analytical merger rate model which makes minimal assumptions about the cosmological history of MBHB evolution and can capture the key characteristics of simulation results to investigate the astrophysical implications of current (Shannon et al., 2013; Lentati et al., 2015; Arzoumanian et al., 2016), and future plausible (Siemens et al., 2013; Ravi et al., 2015) PTA results (either an upper-limit or a detection). Because our model is fully general – not committing to any particular cosmological MBHB merger history – we can identify and separate features of the merger history that are constrained by PTA data alone, from those that can only be constrained by adopting a particular merger history (*e.g.* as done in Shannon et al., 2013; Arzoumanian et al., 2016) – in other words, by applying a particular cosmological prior. Because our model is capable of reproducing the MBHB cosmic population found in cosmological simulations for certain choices of parameters, our results will be consistent with (but much broader than) those that would be obtained under a choice of specific classes of MBHB merger history models.

In section 2.2 we describe our method and merger rate model. In section 2.3 we present our results for several upper limits and for a possible future detection, and we discuss their implications for the population of MBHBs. We present our conclusions in section 2.4.

2.2 Model and method

2.2.1 Astrophysical model

For the standard scenario of circular binaries driven by radiation reaction only, the characteristic strain of the GWSB, h_c at frequency f is (Phinney, 2001):

$$h_c^2(f) = \frac{4G^{5/3}}{3\pi^{1/3}c^2}\, f^{-4/3} \int \mathrm{d}\log_{10}\mathscr{M} \int \mathrm{d}z\, (1+z)^{-1/3}\, \mathscr{M}^{5/3} \frac{\mathrm{d}^3N}{\mathrm{d}V_\mathrm{c}\mathrm{d}z\mathrm{d}\log_{10}\mathscr{M}}\,, \quad (2.1)$$

where z is the redshift and $\mathscr{M}$ is the chirp mass related to the binary component masses (m_1, m_2) by equation 1.10. The integral sums over the sources in z and $\mathscr{M}$ weighted by the distribution of the source population, $\mathrm{d}^3N/\mathrm{d}V_\mathrm{c}\mathrm{d}z\mathrm{d}\log_{10}\mathscr{M}$, the number of binary mergers per co-moving volume, redshift and (rest-frame) chirp mass interval. Equation 2.1 follows directly from equation 1.28, where here we have made the choice of $N(z, \log_{10}\mathscr{M}) = \mathrm{d}^3N/\mathrm{d}V_\mathrm{c}\mathrm{d}z\mathrm{d}\log_{10}\mathscr{M}$ to describe our population. We choose a simple model for this, described by

$$\begin{aligned}\frac{\mathrm{d}^3N}{\mathrm{d}V_\mathrm{c}\mathrm{d}z\mathrm{d}\log_{10}\mathscr{M}} \;=\; \dot{n}_0 &\left[\left(\frac{\mathscr{M}}{10^7\,\mathrm{M}_\odot}\right)^{-\alpha} \exp\left(-\frac{\mathscr{M}}{\mathscr{M}_*}\right)\right] \\ &\left[(1+z)^{\beta} \exp\left(-\frac{z}{z_0}\right)\right] \frac{\mathrm{d}t_r}{\mathrm{d}z}\,, \end{aligned} \quad (2.2)$$

where t_r is the time in the source rest-frame and $\mathrm{d}t_r/\mathrm{d}z$ is given by equations 1.21 and 1.22. Throughout we use $H_0 = 70\,\mathrm{km\ s^{-1}\ Mpc^{-1}}$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\Omega_k = 0$. Following general astrophysical assumptions, we consider a scenario where the GW background is produced by MBHBs in the redshift and chirp mass range of $0 \leq z \leq 5$ and $10^6 \leq \mathscr{M}/\mathrm{M}_\odot \leq 10^{11}$. These ranges set the integration limits of equation 2.1.

The model shown in equation 2.2 is described by five parameters. The parameter $\dot{n}_0$ is the normalised merger rate per unit rest-frame time, co-moving volume and logarithmic

(rest-frame) chirp mass interval. The parameters β and z_0 describe the distribution of the sources in redshift: β controls the low-redshift power-law slope, and z_0 is the high-redshift cut-off for the distribution; the peak of the merger rate $\mathrm{d}^2N/\mathrm{d}t_r\mathrm{d}V_\mathrm{c}$, corresponds to a redshift $(z_0\beta - 1)$. The parameters α and $\mathscr{M}_*$ provide a similar description of the chirp mass distribution. The model was chosen to capture the expected qualitative features of the cosmic MBH merger rate without restricting to any particular merger history; for example, it can reproduce rates extracted from merger tree models (Volonteri et al., 2003; Sesana et al., 2008), and large scale cosmological simulations of structure formation (Springel et al., 2005; Sesana et al., 2009).

The characteristic amplitude has a simple power-law scaling, and we can re-write equation 2.1 as

$$h_c(f) = A_{1\mathrm{yr}} \left(\frac{f}{f_{1\mathrm{yr}}} \right)^{-2/3}, \tag{2.3}$$

where $A_{1\mathrm{yr}}$ is the characteristic amplitude at the reference frequency $f_{1\mathrm{yr}} = 1\,\mathrm{yr}^{-1}$, which is customarily used when quoting limits in the PTA literature. A single number, the amplitude $A_{1\mathrm{yr}}$, carries the whole information about the merging history of MBHBs (within the model considered in this chapter), that one wishes to reconstruct from the observations.

2.2.2 Method

The objective is to put constraints on the population parameters, which we denote by θ, given the results of PTA analyses. In our case θ, is a 5-dimensional parameter space, $\theta = \{\dot{n}_0, \beta, z_0, \alpha, \mathscr{M}_*\}$. We want to compute the posterior density function (PDF) of the parameters given PTA observations, denoted by d. The population parameters fully specify the gravitational-wave signal $h_c(f;\theta)$ (equations 2.1 and 2.3), which in turn specifies the statistical properties of the GW-induced deviations to pulse arrival times –

the PTA observable. Given data from pulsar timing and our model for the merger rate (equation 2.2), we use Bayes theorem to find the posterior distribution of the model parameters which for this case we write as (see also section 1.8),

$$p(\theta|d) = \frac{p(\theta)p(d|A_{1\mathrm{yr}}(\theta))}{p(d)}, \tag{2.4}$$

where $p(d|A_{1\mathrm{yr}}(\theta))$ is the PTA likelihood for a given GWSB, $h_c(f;\theta)$, $p(\theta)$ is the prior on the model parameters and $p(d)$ is the evidence. In standard analysis of the PTA data, constraints are put on the GW characteristic amplitude at periods of one year, $A_{1\mathrm{yr}}$, which in turn is a function of the parameters of the underlying population, specified by $h_c(f;\theta)$. The PTA analysis uses a likelihood function, $p(d|A_{1\mathrm{yr}}(\theta))$, which we approximate as described below. Our method does not rely on this approximation; we use it only for analytical convenience in this work. If a given PTA analysis provides a posterior distribution for $A_{1\mathrm{yr}}$ then a straightforward re-weighting can produce the corresponding likelihood required for our analysis (if flat priors on $A_{1\mathrm{yr}}$ are used in the analysis then the re-weighting is trivial because the posterior and the likelihood are proportional to each other).

In this work we consider the two cases in which the PTA analysis provides either an upper-limit or a detection. For the upper-limit scenario we model $p(d|A_{1\mathrm{yr}})$ using a Fermi-like distribution:

$$p_{\mathrm{ul}}(d|A_{1\mathrm{yr}}) \propto \frac{1}{\exp\left(\frac{A_{1\mathrm{yr}}-A_{\mathrm{ul}}}{\sigma_{\mathrm{ul}}}\right)+1}, \tag{2.5}$$

where A_{ul} is the upper-limit value returned by the actual analysis and the sharpness of the tail-off, σ_{ul} can be adjusted to give an upper limit with a chosen confidence, which

we set at 95%. We model a detection scenario using a Gaussian in the logarithm of $A_{1\mathrm{yr}}$:

$$p_{\mathrm{det}}\left(d|A_{1\mathrm{yr}}\right) \propto \exp\left(-\frac{\left(\log_{10}(A_{1\mathrm{yr}})-\log_{10}(A_{\mathrm{det}})\right)^2}{2\sigma_{\mathrm{det}}^2}\right), \tag{2.6}$$

at a chosen level of detection, A_{det}. We make the choice for the width of the detection to be $\sigma_{\mathrm{det}} = 0.2$ for this study. This value is chosen to provide what we feel to be a reasonable width for a detection scenario which is not so wide as to be consistent with a non-detection, however the exact results of our analysis would be altered by a different choice. We compute the marginalised distribution on the model parameters θ using two independent sampling techniques (see also section 1.8.2), to verify the results of our analysis: a nested sampling approach (Veitch and Vecchio, 2010) and *emcee*, an ensemble Markov Chain Monte Carlo sampler (Foreman-Mackey et al., 2013).

Our priors on the model parameters are set as follows. We use a prior on $\dot{n}_0$ that is flat in $\log_{10}\dot{n}_0$ down to a lower limit, which we set to $\dot{n}_0 = 10^{-20}$ Mpc^{-3} Gyr^{-1}, after which it is flat in $\dot{n}_0$ to zero. Above 10^{-20} Mpc^{-3} Gyr^{-1}, the prior is flat in $\log_{10}\dot{n}_0$ up to 10^3 Mpc^{-3} Gyr^{-1}. This upper bound on the prior is set by the ultra-conservative assumption that all the matter in the universe is formed by MBHs, giving a maximum for the number of MBHBs per volume and time. Our prior allows for the number of mergers to span many orders of magnitude (flat in log) but avoids divergence as $\dot{n}_0 \to 0$. It also allows for the *absence* of MBHB binaries merging within an Hubble time. The priors for the other parameters are uniform within ranges that incorporate values that give a good fit to semi-analytical merger tree models (see *e.g.* Sesana et al., 2008, 2009; Sesana, 2013b): $\alpha \in [-3.0, 3.0]$, $\beta \in [-2.0, 7.0]$, $z_0 \in [0.2, 5.0]$ and $\log_{10}\mathscr{M}_*/\mathrm{M}_\odot \in [6.0, 9.0]$. While our prior allows for parameter values that can reproduce the merger rates of detailed models, it is uninformative in that we do not assume that the merger rate distribution *must* take values from those models. In this way we do not favour, nor

restrict our analysis to any particular model, allowing for inference to be made with PTA results and minimal assumptions from other observations. Our priors reflect large theoretical uncertainties about MBHB formation and evolution scenarios, and the lack of any confirmed MBHB candidate (see however Graham et al., 2015b; Liu et al., 2015; Graham et al., 2015a).

Our method is summarised as follows:

(i) produce a likelihood for $A_{1\mathrm{yr}}$ (in the case of an actual analysis by using smoothed posterior samples from PTA results, re-weighted if necessary depending on the prior)

(ii) choose a model for the merger rate of MBHBs

(iii) produce posterior density functions for the model parameters from which we can infer properties of the MBHB population.

2.3 Results

Current upper limits on the GWSB obtained recently are $A_{1\mathrm{yr}}^{(95\%)} = 1 \times 10^{-15}$, 1.5×10^{-15}, 3×10^{-15} for the PPTA (Shannon et al., 2015), NANOGrav (Arzoumanian et al., 2016), and the EPTA (Lentati et al., 2015), respectively. The sensitivity gain provided by the addition of new pulsars to the PTAs and more recent data sets may allow in the short-to-mid term to reach a sensitivity below $A_{1\mathrm{yr}} = 1.0 \times 10^{-15}$, and in the more distant future $A_{1\mathrm{yr}} \sim 10^{-16}$ (Siemens et al., 2013; Ravi et al., 2015). As a consequence, here we consider three PTA analysis outcomes: (i) an upper-limit at 95% confidence of 1×10^{-15}, which represents the present state of play and either (ii) an upper-limit (at 95% confidence) of 1×10^{-16} or (iii) a detection at the same level, that is $A_{\mathrm{det}} = 1 \times 10^{-16}$, and $\sigma_{\mathrm{det}} = 0.2$ in equation 2.6, which describes possible results

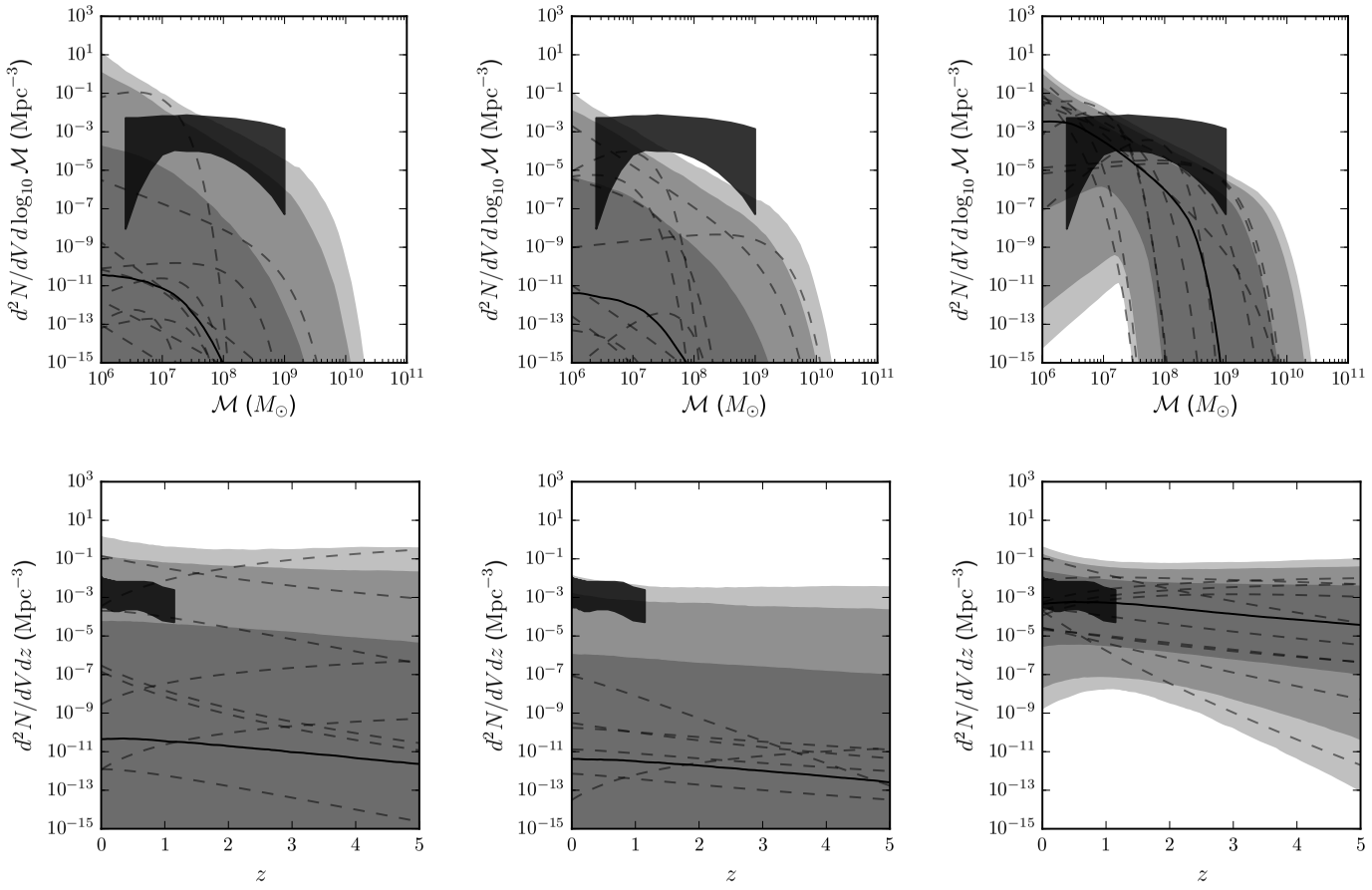


Figure 2.1: Posteriors for the merger rate density. The top row shows the merger rate density in chirp mass (integrated over redshift), $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M}$ and the bottom row in redshift (integrated over chirp mass), $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}z$, for two 95% confidence upper limits at 1×10^{-15} (left) and 1×10^{-16} (centre) and a detection at 1×10^{-16} (right), as described in the text. We consider contributions to the gravitational wave background from massive black holes in the chirp mass range $10^6 \leq \mathscr{M}/\mathrm{M}_\odot \leq 10^{11}$ and redshift range $0 \leq z \leq 5$. The solid black line gives the posterior median; dark grey, mid-grey and light-grey bands show the central 68%, 95%, and 99% credible interval, respectively. The dashed lines show random draws from the posterior. For comparison, the overlaid dark areas represent the 99.7% confidence regions predicted by the MBH assembly models of Sesana (2013b). For these models, we show only the chirp mass in the range $\approx 10^6\ \mathrm{M}_\odot - 10^9\ \mathrm{M}_\odot$ as outside this interval the lower percentile is zero. For the redshift range, these models only consider MBHB mergers for $z \lesssim 1.3$. Reproduced from Middleton et al. (2016).

coming from the expected improvements of the PTA sensitivity in the next five-to-ten years.

The main results of our analysis are summarised in figure 2.1, which shows the inferred posterior distribution of the merger history of MBHBs in terms of their co-moving volume merger density per redshift and per (logarithm of) chirp mass intervals. Figure 2.2 provides PDFs on selected parameters based on current PTA limits, and figure 2.3 provides a similar summary for a future limit or a detection at the level described above.

We consider first the implications of current limits. The PDFs on the parameters $\dot{n}_0$ and $\mathscr{M}_*$ of the model are shown in figure 2.2. We do not provide the equivalent plots for α, β and z_0 as they are equivalent to the prior. Figure 2.2 clearly shows that the present PTA limits enable us to reduce the allowed normalisation of the MBHB merger rate density to $\dot{n}_0 \lesssim 5 \times 10^{-3}$ Mpc^{-3} Gyr^{-1} with 95% confidence, but yield no additional constraints on the other parameters of the model.

Our model contains parameters describing the shape of the merger rate distribution in redshift and chirp mass. The PDFs of those parameters induce a posterior density on $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M}$ and $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}z$, integrating over redshift and chirp mass respectively, and these are shown in figure 2.1. We see that current observations limit the maximum merger density as a function of mass, but place no constraints on the shape of the distribution. The number of sources per frequency bin contributing to the signal can be computed using equations 2.2 and 1.15 followed by integration over the redshift. From this, we see that $\mathrm{d}^2N/\mathrm{d}f\mathrm{d}\log_{10}\mathscr{M} \propto f^{-11/3}$ and find that for masses above a few $\times 10^9$ M$_\odot$, our upper limit on $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M}$ implies that at a frequency around 1.8 nHz there is fewer than one source per frequency bin (taken to be $\Delta f = 1/T$, with $T = 17.66$ yr, the timespan of current EPTA datasets (Lentati et al., 2015). This means that at those large masses, the assumption that the observed GW signal is stochastic

is violated, and our analysis cannot be used to constrain the exact shape of the mass function here (in this case a different PTA search approach would be necessary; see *e.g.* Babak et al., 2015; Arzoumanian et al., 2014; Taylor et al., 2015). While current PTA observations provide feeble constraints on the shape of the mass distribution, they yield no information about the redshift distribution; the bottom-left panel of figure 2.1 shows no structure in $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}z$.

It is useful to compare these results to limits on the MBHB merger rates implied by binary candidates reported in the literature and to specific theoretical models. Let us first consider what is known *observationally* today. A few MBHB candidates have been reported recently. Graham et al. (2015b) suggested the possible observation of a MBHB at redshift $z = 0.2784$ with (rest frame) total mass $\log(M/\mathrm{M}_\odot) \sim 8.5$ and period of ~ 1884 days. Liu et al. (2015) reported the observation of a potential MBHB at $z = 2.060$ with a shorter period of 542 days and primary MBHB mass $\log(M/\mathrm{M}_\odot) \sim 9.97$. Using the redshift to calculate the enclosed volume and the binary parameters for the time to merger we can estimate the predicted rate from each of these observations. Assuming that these two systems are indeed MBHBs, and that their constituents are of comparable mass, they imply merger rates of $\approx 3 \times 10^{-7}$ Mpc^{-3} Gyr^{-1} and ≈ 0.01 Mpc^{-3} Gyr^{-1}. This latter number is much higher as it takes into account that the source has been found in an analysis of 10 of the Pan-STARRS1 Medium Deep Survey fields each of 8 deg^2, however the numbers we calculate are an upper limit based on extrapolation over the sky and would be significantly lower if no more MBHBs are found in the rest of the survey. In turn these values yield a merger density $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M} \approx 10^{-6}$ Mpc^{-3} at $\mathscr{M} \approx 3 \times 10^8$ M$_\odot$ and ≈ 0.1 Mpc^{-3} at $\mathscr{M} \approx 10^{10}$ M$_\odot$, respectively. The upper left panel of figure 2.1 clearly shows that the rate density inferred from Graham et al. (2015b) is consistent with current upper limits, while that inferred from Liu et al. (2015) is well above the 99% credible interval implied by current PTA results. Based on our

assumption that this source was found in such a small survey, is therefore unlikely that this source is a MBHB with the claimed parameters. Other proposed MBHBs in the literature (Valtonen et al., 2012; Kun et al., 2014) imply merger density estimates of 5×10^{-5} Mpc^{-3} at $\mathscr{M} \approx 10^{9}$ M$_{\odot}$ and 3×10^{-6} Mpc^{-3} at $\mathscr{M} \approx 3.5 \times 10^{8}$ M$_{\odot}$, which are consistent with the current upper limits.

On the theoretical side, current limits are consistent with the assumption that most Milky-Way-like galaxies contain a MBH in the mass range considered here that undergoes ~ 1 major merger in an Hubble time. The density of Milky-Way-like galaxies is 10^{-2} Mpc^{-3}, which yields an estimate of $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M} \sim 10^{-3}$ Mpc^{-3}, which is consistent with our results at $\mathscr{M} \sim 10^{6}$ M$_{\odot}$, appropriate for a typical MBHB forming in the merger of Milky-Way-like galaxies. We also compare the limits on $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}\log_{10}\mathscr{M}$ and $\mathrm{d}^2N/\mathrm{d}V_\mathrm{c}\mathrm{d}z$ with specific distributions obtained from predictions of astrophysical models for the cosmic assembly of MBHs. We consider the models presented in Sesana (2013b), extended to include the most recent MBH-galaxy scaling relations (Kormendy and Ho, 2013). These models produce a central 99% interval of $A_{1\mathrm{yr}} \in [2 \times 10^{-16}, 4 \times 10^{-15}]$. The 99.7% confidence region in the merger density from those models is marked by a dark-shaded area in each panel of figure 2.1. Two conclusions can be drawn: (i) present MBHB population models are consistent with current PTA limits; (ii) those models are drawn from a very restricted prior range of the parameters that control the evolution of MBHBs, driven by specific assumptions on their assembly history. For example, in those models there is a one-to-one correspondence between galaxy and MBH mergers. Our results are consistent with the conclusions drawn by Shannon et al. (2013) about the implications of the PPTA limit for the MBHB merger history. However, since they consider specific models that lie close to the upper end of the 99% credible range allowed by current limits, they emphasise the fact that PTA limits might soon be in tension with those specific classes of models.

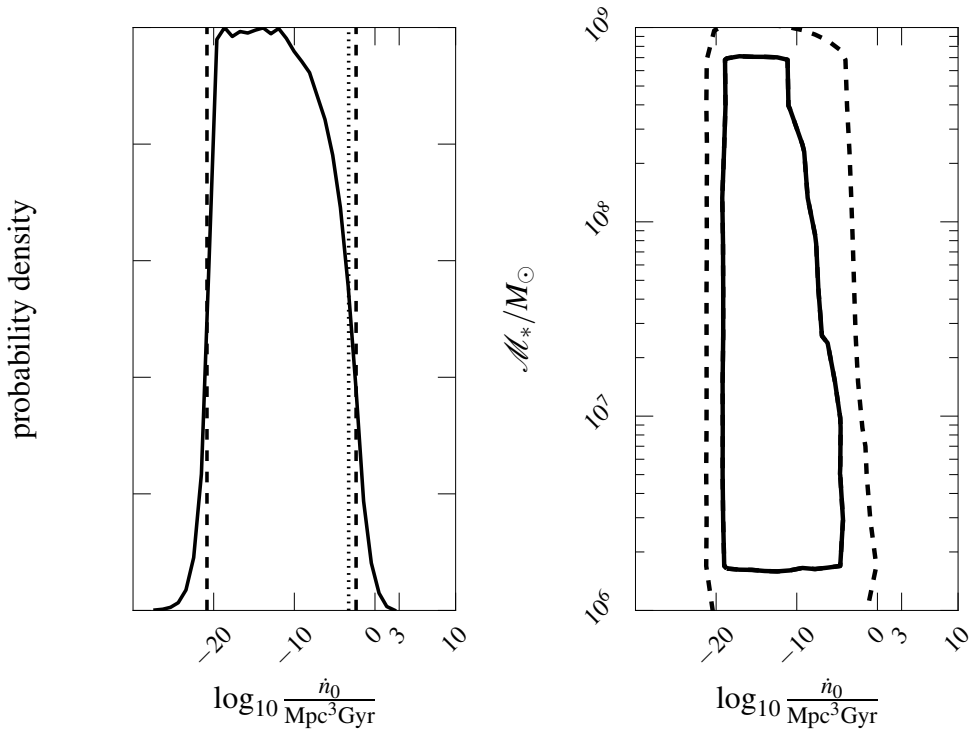


Figure 2.2: Marginalised posterior distributions for selected astrophysical parameters for the case of a 95% upper-limit of 1×10^{-15}, corresponding to the current status of observations. Left panel: The marginalised posterior density function for the merger rate parameter, $\dot{n}_0$. The dashed lines mark the 95% confidence width ($-20.8 \leq \log_{10}[\ \dot{n}_0/\ \mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}] \leq -2.3$) while the dotted line marks the 95% upper limit ($\log_{10}[\ \dot{n}_0/\ \mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}] \leq -3.3$). Right panel: the two-dimensional marginalised posterior density funtion on $(\mathscr{M}_*, \dot{n}_0)$. The contours mark the 67% (solid) and 95% (dashed) confidence regions. In both panels, the left hand side of the distribution in $\log_{10}(\ \dot{n}_0/\ \mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1})$ follows our prior, while the right hand side is determined by the PTA upper limit. Reproduced from Middleton et al. (2016)

We turn now to consider what we could infer about the MBHB merger history in the future as PTA sensitivity increases. To be definitive, we consider both an upper limit and a detection at the level of $A_{1\mathrm{yr}} = 10^{-16}$. Selected marginalised PDFs on the model parameters are shown in figure 2.3, where we see a slight correlation in the 2-D marginalised PDF of $(\mathscr{M}_*, \dot{n}_0)$, as expected. This is simply explained by considering the Schechter-like mass profile of equation 2.2: as the characteristic mass $\mathscr{M}_*$ decreases, and therefore the exponential cut-off of MBHB progressively depletes the high-mass portion of the population, a given value of the GW characteristic amplitude allows for a larger overall normalisation, $\dot{n}_0$. The posterior MBHB merger densities per logarithm of chirp mass and redshift are shown in figure 2.1. For the case of an upper-limit the results are qualitatively similar to the case of the present PTA upper-limit, simply scaled accordingly. In particular, despite the much tighter limit on the overall merger rate we are still unable to place any meaningful constraint on the redshift distribution of merging MBHBs. The overall merger density as a function of redshift shifts by two orders of magnitude and the same is true for the merger density as a function of mass. Note that a non-detection at this level might pose a serious challenge to currently favoured theoretical MBH assembly models with simple black hole dynamics, as shown in the upper-centre panel of figure 2.1.

In the case of detection the posterior on the shapes of the merger rate distribution in redshift and chirp mass are plotted on the right panels of figure 2.1. We still obtain essentially no bounds on the shape of the merger rate density in redshift. We also obtain no meaningful lower bound on the merger rate density for chirp masses. That is, there is no chirp mass at which we can bound the merger density above a rate physically indistinguishable from zero; we know that some MBHBs merge, but we cannot determine *which ones*. Additional information, such as theoretical assumptions, electromagnetic observations constraining the mass spectrum of merging black holes

(like those discussed earlier in this section), or gravitational wave observations that measure the binary mass spectrum directly (such as those of an eLISA-like instrument; Amaro-Seoane et al., 2013), are required to place any constraints on the masses of the merging systems. For example, if we accept the priors provided by Sesana (2013b), the mass function of merging MBHBs can be determined more precisely, as shown by the overlap between our posterior and the dark band in the upper–right panel of figure 2.1.

2.4 Conclusions

We have considered the implications of current PTA limits on the GWSB to constrain the merger history of MBHBs. Using a general model for the mass and redshift evolution of MBHBs in circular orbits driven by radiation reaction, we find that existing PTA results alone place essentially no constraints on the merger history of MBHBs. We also find that even with an increase in amplitude sensitivity of an order of magnitude, and assuming that a detection is made, no bounds can be put on the functional form of the merger rate density in redshift and chirp mass unless additional information coming from a different set of observations is available.

Finally, we want to caution the reader that the results presented here apply only within the model assumptions that have been made. We have considered a generic (and well justified) functional form for the MBHB merger rate density, but if one chooses a significantly different form (and associated priors for the parameters), results could be different (even radically so). Moreover, it has been suggested that physical effects other than radiation reaction, such as gas and/or interactions with stars (*e.g.* Kocsis and Sesana, 2011; Sesana, 2013a; Sampson et al., 2015), could affect the evolution of MBHBs. These effects are not included in our model, and their impact on astrophysical inference needs to be evaluated in the future.

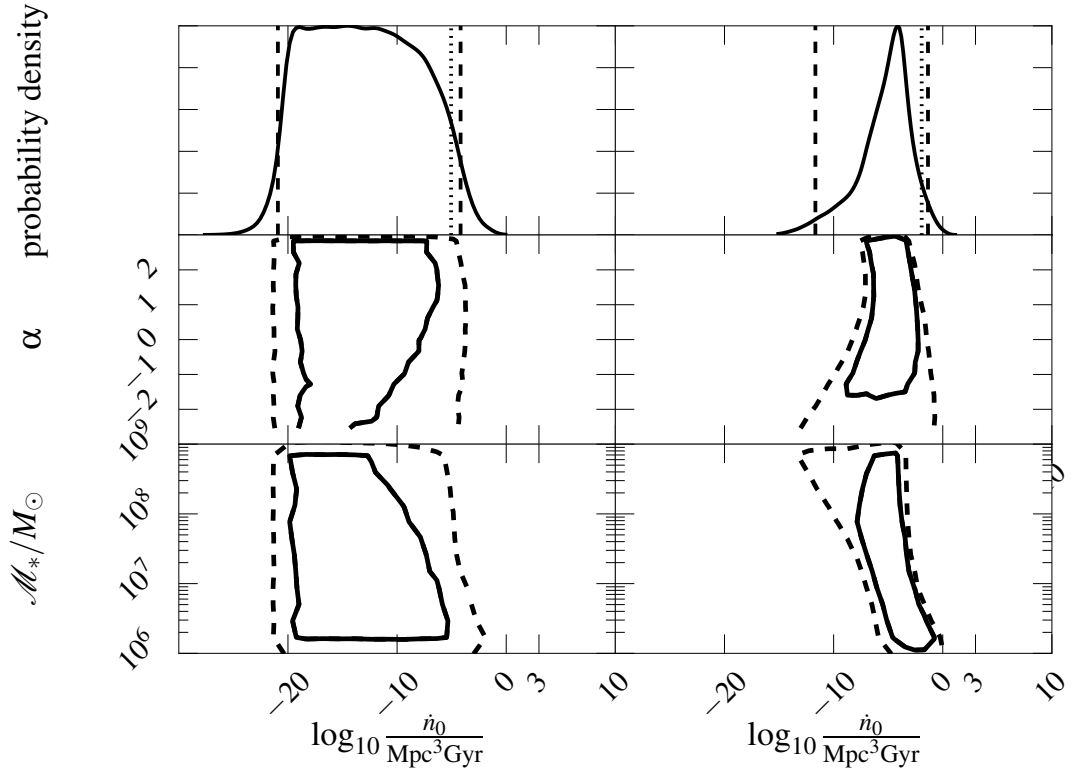


Figure 2.3: Posterior distribution for the upper limit (left) and detection (right) at $A_{1\mathrm{yr}} = 1 \times 10^{-16}$. The top panels show the one dimensional posterior distribution for the merger rate parameter, $\dot{n}_0$. The dashed lines mark the 95% confidence width (upper limit: $-20.9 \leq \log_{10}(\dot{n}_0/\mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}) \leq -4.2$; detection: $-11.7 \leq \log_{10}(\dot{n}_0/\ \mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}) \leq -1.3$) and the dotted line the 95% upper limit (upper limit: $\log_{10}(\dot{n}_0/\mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}) = -5.0$; detection: $\log_{10}(\dot{n}_0\ \mathrm{Mpc}^{-3}\ \mathrm{Gyr}^{-1}) = -1.9$). The central and bottom panels show the two dimensional posterior distributions for $\dot{n}_0$ with the mass parameters α and $\mathscr{M}_*$. The solid and dashed contours mark the 67% and 95% confidence regions, respectively. Reproduced from Middleton et al. (2016)

Chapter 3

No cause for concern: applying our method to current upper limit results

This chapter is split into two halves, the first being a review of work from Chen et al. (2017a,b) and Chen et al. (2017c) relevant to the second half, which reproduces verbatim the paper Middleton et al. (2017), currently under consideration by Nature Communications. The paper Middleton et al. (2017) describes the application of a method similar to that detailed in chapter 2 to the most recent upper limits. However, as the project relies heavily on improvements to the method of chapter 2, which were introduced by Chen et al. (2017a,b) and Chen et al. (2017c), we provide here an introduction covering details from those works which are relevant to Middleton et al. (2017). The full transcript of Chen et al. (2017a,b) (of which I am also a contributing author) is also reproduced in appendix A.

3.1 Background

Here we provide details of two aspects of the work carried out by Chen et al. (2017c) and Chen et al. (2017a,b) which are relevant to this chapter. The first is the model used to describe an eccentric population of merging massive black hole binaries (section 3.1.1) and the other is construction of the likelihood for a frequency-dependent upper limit on the gravitational wave stochastic background (section 3.1.2).

3.1.1 Eccentric massive black hole binaries gravitational wave background

In chapter 2, we covered how upper limits and future detections of the GWSB from the merging population of MBHBs can be used to infer limited information on the properties of the source population. One assumption made in the work covered in chapter 2 is that the binaries follow circular orbits which are driven by radiation reaction only. This assumption means that the gravitational wave spectrum is described by a power law with characteristic strain $h_c \propto f^{-2/3}$ as shown by equation 1.28 and an upper limit placed at a single frequency ($f_{yr} = \mathrm{yr}^{-1}$ in chapter 2) is enough to describe the entire spectrum. However, if eccentricity is allowed, the spectrum will be more complex than the simple power law and it becomes more advantageous to make full use of upper limits (or detections) at multiple frequency bins.

Here we follow the results of Chen et al. (2017c) (which we label CSD17 from here on), who cover two aspects of binary evolution which can lead to a departure from the $f^{-2/3}$ spectrum. One is the effect of eccentricity of an individual binary spectrum and the other is the effect of the depletion of source numbers at high frequencies, leading to a drop in the characteristic amplitude spectrum.

After a galaxy merger, if a MBHB is to reach a separation at which GW emission is

the dominant process in its evolution, then the black holes must reduce their separation via interaction with the environment (see also section 1.5.1). The effect of the environment is twofold; on the one hand, these interactions remove energy from the system, shrinking the binary's semi-major axis a as (Quinlan, 1996; Sesana and Khan, 2015)

$$\frac{da}{dt} = \frac{a^2 G\rho}{\sigma} H, \tag{3.1}$$

where ρ is the density of stars in the vicinity of the MBHB, σ is the stellar velocity dispersion and $H \sim 15 - 20$ is a numerical coefficient. The other effect is that during these interactions the binary can increase in eccentricity. Both of these effects can lead to binaries sweeping through the lower frequencies more quickly, leading to a possible turnover in the spectrum at lower frequencies as shown by figure 3.1. Interactions can be with stars, gas or another massive black hole (see section 1.5.1). The effect considered by CSD17 is from stellar hardening, which may leave the binary with some eccentricity at the time when it decouples from the environment, from which point they will continue to evolve via GW radiation. They show that this decoupling occurs below the PTA sensitivity window and the main effect from the environment on the spectrum in-band is due to the eccentricity that the binaries are left with at decoupling (see fig. 6 of CSD17).

Here we summarise the key results of CSD17, relevant to Middleton et al. (2017) illustrating the changes made to the model from Middleton et al. (2016). Starting with the result from Phinney (2001) as described in section 1.5.3, the characteristic strain is (see also equations 1.26 and 1.28)

$$h_c^2(f) = \frac{4G}{\pi c^2 f} \int_0^\infty \mathrm{d}z \int_0^\infty \mathrm{d}\mathscr{M} \frac{\mathrm{d}^2 n}{\mathrm{d}z\, \mathrm{d}\mathscr{M}} \frac{\mathrm{d}E}{\mathrm{d}f_r}. \tag{3.2}$$

However, what differs when allowing for eccentric binaries is the final derivative (see

also equations 3 and 4 of CSD17),

$$\frac{\mathrm{d}E}{\mathrm{d}f_r} = \sum_{n=1}^{\infty} \frac{1}{n} \frac{\mathrm{d}E_n}{\mathrm{d}t} \frac{\mathrm{d}t}{\mathrm{d}e_n} \frac{\mathrm{d}e_n}{\mathrm{d}f_n}, \tag{3.3}$$

where $f_n = f_r/n$ is the orbital rest frame frequency of the binary (the nth harmonic has observed frequency f and e_n is the eccentricity at that orbital frequency), $\mathrm{d}E_n/\mathrm{d}t$ is the luminosity of the nth GW harmonic (for full details see Peters and Mathews, 1963, CSD17 and references therein). The parameter of interest for this work is the transition eccentricity e_t at which the binaries decouple from the environment and their evolution becomes dominated by radiation reaction. In order to produce an efficient means of computing the GWSB, CSD17 use the fact that a binary with a given frequency and eccentricity produces an identical spectrum to one at some earlier stage of evolution with lower frequency and higher eccentricity once the correct transformation along the $f^{-2/3}$ diagonal has been made (Huerta et al., 2015; Enoki and Nagashima, 2007). This can be seen visually from figure 3.1 (and is better illustrated by fig. 2 of CSD17). The computation is sped up with the use of a fitting formula for a reference MBHB spectrum which can then be transformed in frequency. The spectrum is then given by

$$h_c^2 = \int \mathrm{d}z \int \mathrm{d}\mathscr{M} \frac{\mathrm{d}^2 n}{\mathrm{d}z\,\mathrm{d}\mathscr{M}} h_{c,\mathrm{fit}}^2 \left(f \frac{f_{p,0}}{f_{p,t}} \right) \left(\frac{f_{p,t}}{f_{p,0}} \right)^{-4/3} \left(\frac{\mathscr{M}}{\mathscr{M}_0} \right)^{5/3} \left(\frac{1+z}{1+z_0} \right)^{-1/3} \tag{3.4}$$

where $h_{c,\mathrm{fit}}$ is a fit to the spectrum for a reference binary with chirp mass $\mathscr{M}_0$ at redshift z_0 and some eccentricity e_0 at a decoupling frequency of f_0 (which defines the peak in the spectrum $f_{p,0}$). In this case we have six parameters describing the spectrum; the five parameters described in chapter 2 ($\dot{n}_0$,β,z_0,α,$\mathscr{M}_*$) and one eccentricity at transition parameter e_t, for which the simplifying assumption is made that all binaries in the population have the same eccentricity at transition. Although this may not be realistic, we use this assumption for our initial study as an extention to the circular binary case,

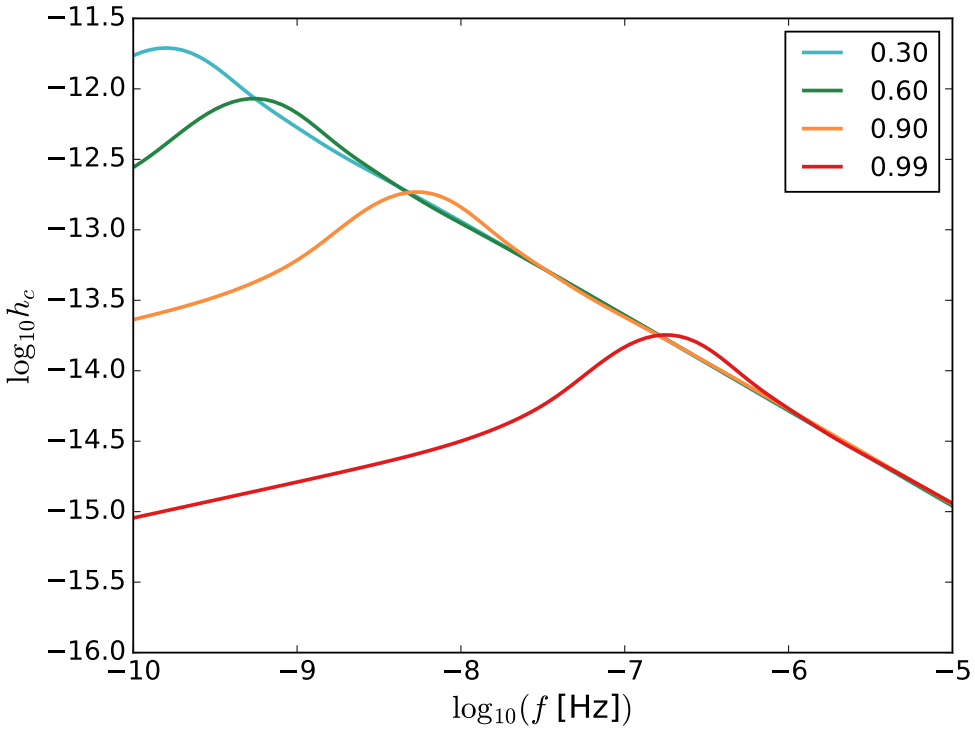


Figure 3.1: The characteristic amplitude spectrum for a fiducial binary used in Chen et al. (2017c) with chirp mass of 4.16×10^8 $M_\odot$, redshift 0.02, reference orbital frequency of 0.1 nHz and different initial eccentricities $\{0.3, 0.6, 0.9, 0.99\}$ as shown by the four lines. The binaries have already entered the regime where GW radiation dominates. Over time binaries will circularize, as can be seen at higher frequencies where all of the spectra follow the same $f^{-2/3}$ power law. Reproduced from Chen et al. (2017c).

which we can further build on in future investigations.

The other effect they consider takes into account the number of sources from the population in each frequency bin. The GWSB is constructed by summing up the contributions of all sources within a frequency bin. For a realistic population of MBHBs, the number of sources decreases at higher masses. This leads to the number of sources at higher frequencies being depleted as shown by Sesana et al. (2008), and some bins may contain only 'fractions' of sources. In their fitting formula, CSD17 take account of this by implementing an upper limit on the mass integral, imposing that there must be at least one source per frequency bin for a stochastic background (see equation 40 of their

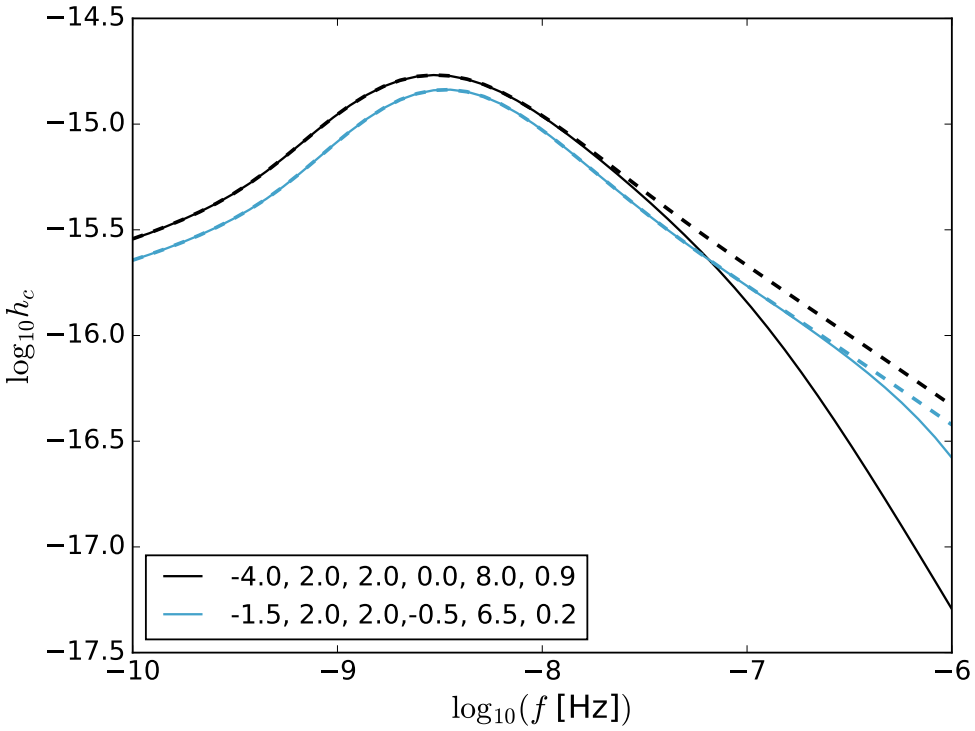


Figure 3.2: The characteristic amplitude for a population of massive black hole binaries using the model described in chapter 2. The numbers listed in the legend indicate the values of the model parameters in the order $\{\log_{10}\dot{n}_0, \beta, z_0, \alpha, \mathscr{M}_*, e_t\}$. For each set of parameters, two lines are displayed; the solid lines display the spectrum including the effect of the high-mass source depletion and the dashed lines show the spectrum without including this effect. Reproduced from Chen et al. (2017c).

work). This effect leads to a depletion of sources at higher frequencies, as can be seen in figure 3.2.

3.1.2 Likelihood for a gravitational wave spectrum

The construction of the likelihood is also more complex when there is a frequency-dependent spectrum. As we would like to be able to take full advantage of the shape of the frequency dependent upper limit, we need to fold the results from each bin into the likelihood calculation. The details of this work can be found in Chen et al. (2017a,b) (and reproduced in appendix A). The aim is to produce a likelihood constructed from

upper limits (or detections which were also considered in Chen et al. (2017a,b)) on a bin-by-bin basis. As in chapter 2, a Fermi-like function is used to represent a single frequency bin upper limit (and Gaussians, for example, could again be used for a bin with a detection). The overall likelihood is constructed by multiplication of the bin-by-bin likelihoods, assuming that they are statistically independent. This method also allows for the combination of detection and upper limit bins as done in Chen et al. (2017a,b).

The following pages reproduce Middleton et al. (2017) in reformatted form.

No tension between assembly models of massive black hole binaries and pulsar observations

The following work is reproduced verbatim from a paper currently under consideration by Nature Communications (Middleton et al., 2017). The text and figures are taken directly form the paper, but may still be edited before publication. The work was done in collaboration with the co-authors listed below. I made contributions towards the code and analysis, wrote the initial draft and assisted editing for the paper text and produced some of the figures.

Authorship:

Hannah Middleton, Siyuan Chen, Walter Del Pozzo, Alberto Sesana, Alberto Vecchio.

Abstract

Pulsar timing arrays (PTAs) are presently the only means to search for the gravitational wave stochastic background from super massive black hole binary populations, considered to be within the grasp of current or near-future observations. The stringent upper limit from the Parkes PTA has been interpreted as excluding ($> 90\%$ confidence) the current paradigm of binary assembly through galaxy mergers and hardening via stellar interaction, suggesting evolution is accelerated or stalled. Using Bayesian hierarchical modelling we consider implications of this upper limit for a range of astrophysical scenarios, without invoking stalling, nor more exotic physical processes. All scenarios are fully consistent with the upper limit, but (weak) bounds on population parameters can be inferred. Recent upwards revisions of the black hole-galaxy bulge mass relation are disfavoured at 1.6σ against lighter models. Once sensitivity improves by an order of magnitude, a non-detection will disfavour the most optimistic scenarios at 3.9σ.

3.2 Introduction

Dedicated timing campaigns of ultra-stable radio pulsars lasting over a decade and carried out with the best radio telescopes around the globe have targeted the isotropic gravitational-wave (GW) background in the frequency region $\sim 10^{-9} - 10^{-7}$ Hz generated by the cosmic population of merging massive black hole binaries (MBHBs). In the hierarchical clustering scenario of galaxy formation, galaxies form through a sequence of mergers (see White and Rees, 1978, and section 1.5.1). In this process, the MBHs hosted at their centre will inevitably form a large number of binaries (Begelman et al., 1980), forming an abundant population of GW sources in the universe. Detecting and/or placing constraints on their emitted signal will thus provide an insight into the formation and evolution of MBHs in connection with their galaxy hosts and will help to better

understand the role played by MBHs in galaxy evolution and the dynamical processes operating during galaxy mergers (see Sesana, 2013a, for a review).

No detection at nHz frequencies has been reported so far. The most stringent constraint on an isotropic background radiation has been obtained through an 11 year-long timing of 4 radio-pulsars by the Parkes Pulsar Timing Array (PPTA). It yields an upper limit on the GW characteristic amplitude of $h_{1\mathrm{yr}} = 1.0 \times 10^{-15}$ (at 95% confidence) at a frequency of 1 yr^{-1} (Shannon et al., 2015). Consistent results, although a factor ≈ 2 less stringent, have also been reported by the European PTA (EPTA; Lentati et al., 2015), the North Amercian Nanohertz Observatory for Gravitational Waves (NANOGrav; Arzoumanian et al., 2016), and the International PTA (IPTA; Verbiest et al., 2016), an international consortium of the three regional PTA collaborations. Those values are in the range of signal amplitudes predicted by state-of-the-art MBHB population models, and can therefore be used to constrain such a population. It has been noted, however, that these limits start to be sensitive to uncertainties in the determination of the solar system ephemeris used in the analysis. Recent unpublished work has in fact found that different ephemeris choices can result in a partial degradation of the upper limit (Hobbs and Dai, 2017). This is still an active area of research which may lead to a small upward revision of the upper limit, a circumstance which, if anything, will strengthen the conclusion of our analysis. Here we consider the most stringent upper limit from the PPTA in order to glean what can be learnt at this stage and also determine whether current MBHB population models are indeed cast into doubt.

Using the PPTA limit, we place bounds on the properties of the sub-parsec population of cosmic MBHBs (in the mass range $\sim 10^7 - 10^{10}$ $\mathrm{M}_\odot$) and explore what constraints, if any, can be put on the salient physical processes that lead to the formation and evolution of these objects. We consider a comprehensive suite of astrophysical models that combine observational constraints on the MBHB population with state-of-the-

art dynamical modelling of binary evolution. The MBHB merger rate is anchored to observational estimates of the host galaxy merger rate by a set of MBH-host relations (see section 3.5 and Sesana, 2013a; Sesana et al., 2016). Rates obtained in this way are well captured by a five parameter analytical function of mass and redshift, once model parameters are restricted to the appropriate prior range (see section 3.5). Individual binaries are assumed to hold a constant eccentricity so long as they evolve via three-body scattering and gradually circularise once GW emission takes over. Their dynamical evolution and emission properties are regulated by the density of the stellar environment (assumed to be a Hernquist profile (Hernquist, 1990) with total mass determined by the MBH mass – galaxy bulge mass relation) and by the eccentricity during the three-body scattering phase, which we take as a free parameter. For each set of model parameters, the characteristic GW strain $h_c(f)$ at the observed frequency f is computed as described in Chen et al. (2017c), and summarised in section 3.5. Our model encapsulates the significant uncertainties in the GW background due to the poorly constrained MBHB merger rate and has the flexibility to produce a low-frequency turnover due to either three-body scattering or high eccentricities. MBHBs are assumed to merge with no significant delay after galaxies merge. As such, the models do not include the effect of stalling or delayed mergers (Simon and Burke-Spolaor, 2016).

We find that although PTAs have well and truly achieved a sensitivity for which detection is possible based on model predictions, the present lack of a detection provides no reason to question these models. We highlight the impact of the MBH-galaxy relation by considering a selection of models which cover the entire range of the predicted background amplitude. To be definitive, we consider four models: an 'optimistic' model (in terms of predicting a high amplitude which is based on Kormendy and Ho, 2013, here labelled KH13), which provides a prediction of the GW background with median amplitude at $f = 1$ yr^{-1} of $h_{1\mathrm{yr}} = 1.5 \times 10^{-15}$; a conservative model (labelled

G09, based on Gültekin et al., 2009), with $h_{1\mathrm{yr}} = 7 \times 10^{-16}$; an ultra-conservative model (labelled S16, based on Shankar et al., 2016), with $h_{1\mathrm{yr}} = 4 \times 10^{-16}$; and finally a model that spans the whole range of predictions within our assumptions (which we label 'ALL'). Note that the latter contains as subsets KH13, G09 and S16, but it is not limited to them. Moreover, model 'ALL' spans, in $h_{1\mathrm{yr}}$, an amplitude range that comfortably includes GW backgrounds estimated by other authors employing different techniques (*e.g.*, McWilliams et al., 2014; Ravi et al., 2015; Kulier et al., 2015; Kelley et al., 2017). Details on the models are provided in section 3.5.

3.3 Results

For each model, we use a Bayesian hierarchical analysis to compute the model evidence (which is the probability of the model given the data and allows for the direct comparison of models) and posterior density functions on the model parameters given the observational results reported by Shannon et al. (2015). We find that the upper limit is now beginning to probe the most optimistic predictions, but all models are so far consistent with the data. Figure 3.3, our main result, compares the predictions under different model assumptions with the observed upper limit. The dotted area shows the prior range of the GW amplitude under the model assumptions, and the orange solid line shows the 95% confidence PPTA upper limit on h_c. The (central) 68% and 90% posterior probability intervals on h_c are shown by the shaded blue bands. The posterior density functions (PDFs) on the right hand side of each plot gives the prior (black-dashed line) and posterior (blue line) for h_c at a reference frequency of $f \sim 1/5$ yr^{-1} (see supplemetary text 2.2 of Shannon et al., 2015).

The difference between the dotted region and the shaded bands in the main panels of figure 3.3 indicates the constraining power of the PPTA limit on astrophysical models

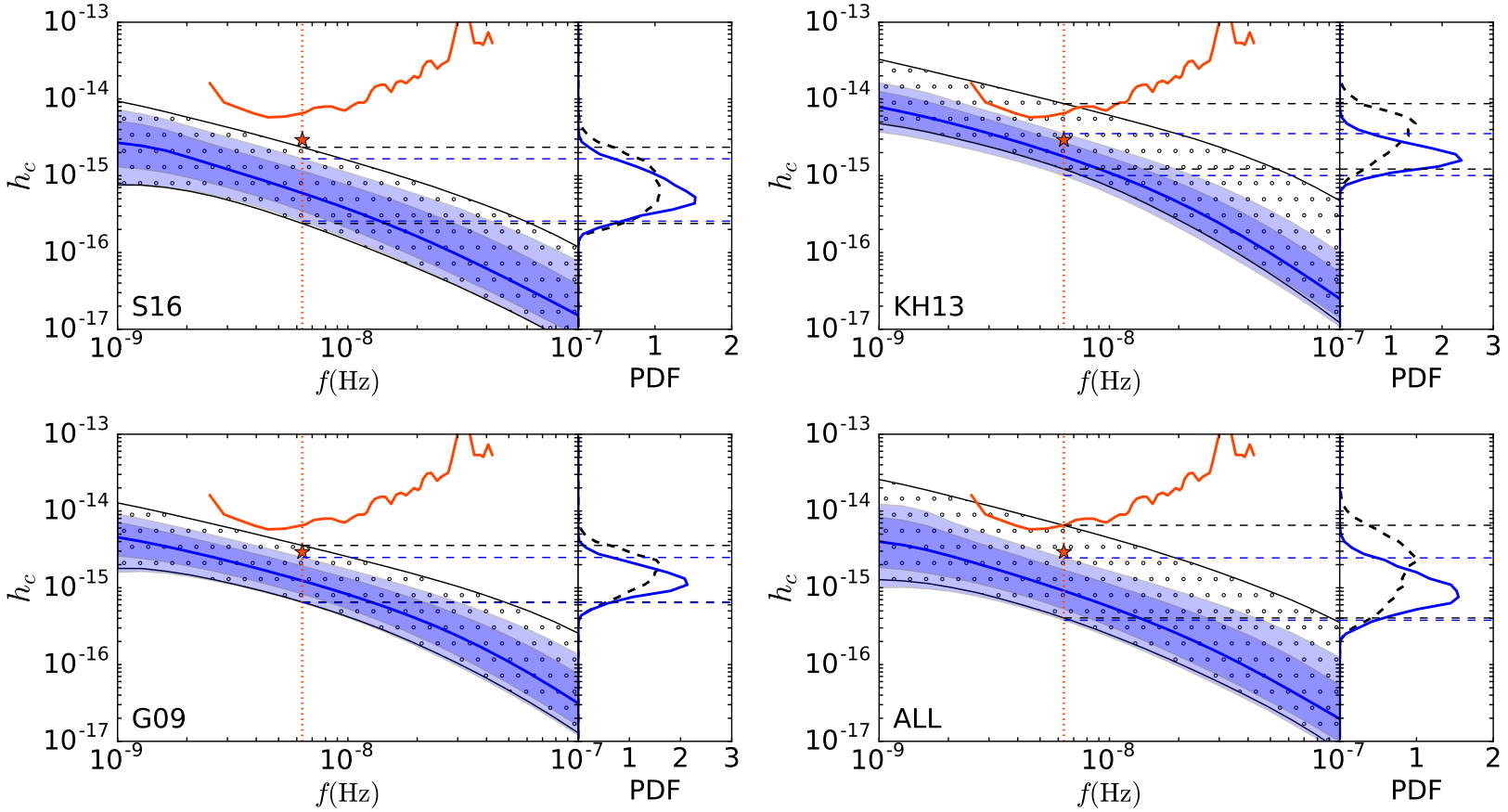


Figure 3.3: The four panels compare the prior and posterior density functions on the GW stochastic background characteristic amplitude in light of the PPTA upper limit for each of the astrophysical models considered here. The central 90% region of the prior is indicated by the dotted band, and the posterior is shown by the progressively lighter blue shading indicating the central 68% and 90% regions, along with the median (solid-blue line). Also shown are the PPTA bin-by-bin limit (orange-solid line) and the corresponding integrated limit assuming $h_c(f) \propto f^{-2/3}$ (red star and red vertical dotted line). The difference in the prior and posterior indicates how much has been learnt from the PPTA data. The right-hand side one-dimensional distributions show the prior (black-dashed) and posterior (blue-solid) at a reference frequency of $f \sim 1/5$ yr^{-1}, with the central 90% regions marked with black-dashed and blue-dashed lines for the prior and posterior respectively. Reproduced from Middleton et al. (2017).

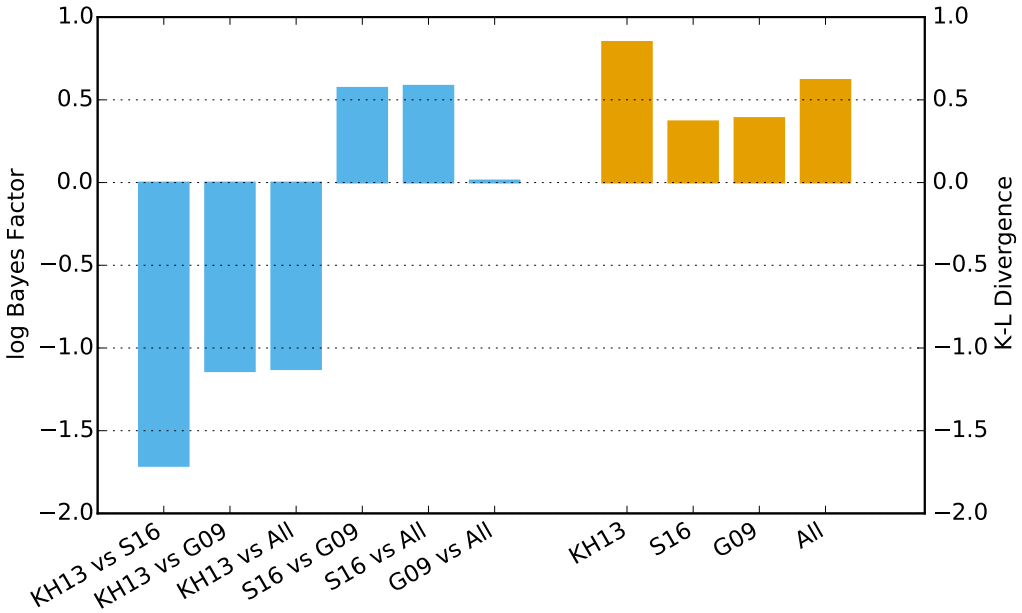


Figure 3.4: We compare the Bayes factors between model pairs (left-hand, blue bars) and the Kullback-Leibler (K-L) divergences between the prior and posterior of characteristic amplitude (right-hand, orange bars). The small range of Bayes factors indicates that there is little to choose from between these models, although KH13 is weakly disfavoured against the others. The K-L divergences also support this conclusion. Although all values are small, KH13 has the largest K-L divergence of the four models indicating the greatest difference between the prior and posterior. Reproduced from Middleton et al. (2017).

– the greater the difference between the two regions, the smaller is the consistency of that particular model with the data. We see that although some upper portion of the allowable prior region is removed from the 90% posterior probability interval (less so for S16), none of the models can be ruled out at any significant level. The confidence bands across the frequency range are constructed by taking the relevant credibility region of the posterior distribution of h_c at each frequency, and therefore the boundaries of each band do not follow any particular functional form as a function of frequency. In addition, although eccentricity is allowed by the data, the power-law spectrum of circular binaries driven by radiation reaction alone can clearly be consistently placed within these bands (see figure 3.6 for further details on the individual parameter posteriors including eccentricity). This can be quantified in terms of the model evidences $\mathcal{Z}$, shown in table 3.1. The normalisation is chosen so that a putative model unaffected by the limit yields $\mathcal{Z} = 1$, and therefore the values can be interpreted as Bayes factors against such a model. None of the posterior probabilities of the models with respect to this putative one show any tension. As an example, for models ALL and S16 we find $e^{-1.23} = 0.3$ and $e^{-0.6} = 0.55$, respectively. Similar conclusions can be drawn from the Kullback-Leibler (K-L) divergences between the prior and posterior on the characteristic amplitude for a given model (with which we measure the difference between the prior and posterior, see also section 1.8.3). For models ALL and S16, these yield 0.62 and 0.37, respectively. As a comparison, these values correspond to the K-L divergence between two Gaussian distributions with the same variance and means approximately 1.1 and 0.8 standard deviations apart for models ALL and S16, respectively[1].

Figure 3.4 summarises the natural logarithm of the ratio of the model evidence, *i.e.* the Bayes factors, between all of the possible combinations of the models, as well

[1] The Kullback-Leibler divergence between two normal distributions $p \sim N(\mu_p, \sigma_p^2)$ and $q \sim N(\mu_q, \sigma_q^2)$ is $\mathrm{D_{KL}}(p||q) = \ln(\sigma_q/\sigma_p) - 1/2 + 1/2\left[(\sigma_p/\sigma_q)^2 + (\mu_p - \mu_q)^2/\sigma_q^2)\right]$. For $\sigma_p = \sigma_q$ and $\mu_p = \mu_q + \sigma_q$ the KL divergence is 0.5.

Model	$h_{1\mathrm{yr}} = 1\times10^{-15}$(PPTA)		$h_{1\mathrm{yr}} = 3\times10^{-16}$		$h_{1\mathrm{yr}} = 1\times10^{-16}$	
	K-L divergence	log$\mathcal{Z}$	K-L divergence	log$\mathcal{Z}$	K-L divergence	log$\mathcal{Z}$
KH13	0.85	-2.36	2.25	-5.68	5.18	-13.17
G09	0.39	-1.2	1.11	-3.35	2.86	-8.26
S16	0.37	-0.6	0.69	-1.62	1.42	-3.82
ALL	0.62	-1.23	1.33	-2.68	2.50	-5.74

Table 3.1: The values in the table show the Kullback-Leibler divergence and natural logarithm of the evidence, log$\mathcal{Z}$, for each of the four astrophysical models given the PPTA upper limit at $h_{1\mathrm{yr}} = 1\times10^{-15}$, and for more stringent limits at the levels of 3×10^{-16} and 1×10^{-16}. Reproduced from Middleton et al. (2017).

as the K-L divergences whose numerical values are listed in table 3.1. Both metrics clearly indicate that there is little to choose from between the models. The least favoured model in the range of those considered here is KH13, with Bayes factors in favour of the others ranging from ≈ 1.13 to ≈ 1.76. These are however values of order unity, and no decisive inference can be made from the data (Kass and Raftery, 1995). Comparisons between each of the individual model parameters (see section 3.5.1) posterior and prior distribution functions are described in figure 3.6 and table 3.2 which further support our conclusions. For KH13, the model that produces the strongest GW background, we find a probability of $e^{-2.36} = 0.094$ with respect to a model that is unaffected by the limit. KH13 is therefore disfavoured at $\sim 1.6\sigma$. This conclusion is reflected in the value of the K-L divergence of 0.85[2]. We note that Shannon et al. (2015) choose in their analysis only a sub-sample of the Sesana (2013a) models, with properties similar to KH13. Our results for KH13 are therefore consistent with the 91%-to-97% 'exclusion' claimed by Shannon et al. (2015).

[2]This is the same K-L between two Gaussian distributions with the same variance and means approximately 1.3 standard deviation apart

3.4 Discussion

It is argued in Shannon et al. (2015) that the Parkes PTA upper limit excludes at high confidence standard models of MBH assembly – *i.e.* those considered in this work – and therefore these models need to be substantially revised to accommodate either accelerated mergers via strong interaction with the environment or inefficient MBHB formation following galaxy mergers. The work presented here does not support either claim. In particular, the posterior parameter distributions (see also figure 3.6) favour neither high eccentricities nor particularly high stellar densities, indicating that a low frequency spectral turnover induced by MBHB dynamics is not required to reconcile the PTA upper limit with existing models. Similar to Taylor et al. (2016), this finding does not support an observing strategy revision in favour of higher cadence observations aimed at improving the high frequency sensitivity, as proposed by Shannon et al. (2015). Likewise, neither stalling nor delays between galaxy and MBHB mergers, which, by construction, are not included in the models considered here, are needed to explain the lack of a detection of GWs at the present sensitivity level. Compared to previous analysis, our work implies a stronger rejection of the statement that there is tension between PTA data and theoretical MBHB population models. For example Simon and Burke-Spolaor (2016) invoked time delays to reconcile the PPTA upper limit with selected MBH-galaxy relations, however they assume a narrow range of possible MBHB merger histories and do not consider MBHB dynamics. The analysis of Arzoumanian et al. (2016) tend to favour a spectral turnover due to either high eccentricity or strong environmental coupling, however they propose a simplified analysis where each relevant physical parameter is accounted for separately. When allowing all the parameters to vary simultaneously, we find that none of them has a critical impact on the inference, and current MBHB population models are broadly consistent with the PTA upper limits

without the need to invoke a low-frequency spectral turnover.

On the other hand, PTA limits are now starting to provide interesting information about the population of merging MBHs. The fact that KH13 is disfavoured at $1.4\,\sigma$ with respect to S16 indicates that the population may have fewer high mass binaries, mildly favouring MBH-host galaxy relations with lower normalisations. This indicates that the gravitational wave background level is likely below the 10^{-15} level, making detection difficult with current telescopes. In this respect, our analysis highlights the importance of upcoming facilities such as MeerKAT (Booth et al., 2009), FAST (Nan et al., 2011) and the Square Kilometer Array (SKA; Dewdney et al., 2009). Their superior timing capabilities, together with their survey potential in finding new stable millisecond pulsars, will provide the necessary ground to improve sensitivity down to $h_{1\mathrm{yr}} \sim 10^{-16}$, which is in line with the lower limit of the expected stochastic gravitational wave background according to our current understanding of MBH evolution (Bonetti et al., 2017b). Although not yet decisive, our findings highlight the potential of PTAs in informing the current debate on the MBH-host galaxy relation. Recent discoveries of over-massive black holes in the brightest cluster ellipticals (McConnell et al., 2011; Hlavacek-Larrondo et al., 2012) led to an upward revision of those relations (McConnell and Ma, 2013; Kormendy and Ho, 2013). However, several authors attribute the high normalisation of the recent MBH-host galaxy relations to selection biases (Shankar et al., 2016) or to the intrinsic difficulty of resolving the MBH fingerprint in measurements based on stellar dynamics (see discussion in Rasskazov and Merritt, 2016). Future facilities such as the Extremely Large Telescope (Gilmozzi and Spyromilio, 2007) and the Thirty Meter Telescope (Sanders, 2013) will likely measure many more MBH masses in elliptical galaxies (Do et al., 2014), providing a better understanding of the MBH-host galaxy relations. PTA limits may therefore be used to gain more information about the other underlying uncertainties in the models, in particular the massive galaxy merger

rate, which is currently poorly constrained observationally (see *e.g.*, Lotz et al., 2011; Mundy et al., 2017).

An important question is what is the sensitivity level required to really put under stress our current understanding of MBHB assembly? If a null result persists in PTA experiments, this will in turn lead to a legitimate re-thinking of the PTA observing strategy to target possibly more promising frequencies of the GW spectrum. To address this question, we simulate future sensitivity improvements by shifting the Parkes PTA sensitivity curve down to provide 95% upper limits of $h_{1\mathrm{yr}}$ at 3×10^{-16} and 1×10^{-16}. The results are summarised in table 3.1 and more details are provided in figure 3.7 and table 3.3). At 3×10^{-16}, possibly within the sensitivity reach of PTAs in the next ≈ 5 years, S16 will be significantly favoured against KH13, with a Bayes factor of $e^{4.06}$, and only marginally favoured over G09, with Bayes factor of $e^{1.76}$. It will still be impossible to reject this model at any reasonable significant level with respect to, say, a model which predicts negligible GW background radiation at $\sim 10^{-9}-10^{-8}$ Hz. However MBH-host galaxy relations with high normalisations will show a $\approx 2\,\sigma$ tension with more conservative models. At 1×10^{-16}, within reach in the next decade with the advent of MeerKAT, FAST and SKA, models KH13, G09 and ALL are disfavoured at $3.9\,\sigma$, $2.5\,\sigma$ and $1.2\,\sigma$, respectively, in comparison to S16. K-L divergences in the range $5.18-1.42$ show that the data are truly informative. S16 is also disfavoured at 2.3σ with respect to a model unaffected by the data, possibly indicating the need of additional physical processes to be included in the models.

3.5 Methods

Here we expand the description of the relevant features of our models and analysis approach. Further details about the astrophysical models can be found in Chen et al.

(2017c) and for the method see Chen et al. (2017a). In section 3.5.1, we present the parametric model describing the GW background generated from a population of eccentric binaries evolving via three-body scattering. In section 3.5.2, we define the astrophysical prior range of the model parameters and anchor them to the empirical estimate of the MBHB merger rate based on observations of close galaxy pairs. In section 3.5.3, we describe the details of the Bayesian hierarchical analysis used in this work.

3.5.1 Analytical description of the GW background

The GW background from a cosmic population of MBHBs is determined by the binary merger rate and by the dynamical properties of the systems during their inspiral. The comoving number density of MBHBs per unit log chirp mass (equation 1.9) and unit redshift, $\mathrm{d}^2n/(\mathrm{d}\log_{10}\mathscr{M}\mathrm{d}z)$, defines the normalisation of the GW spectrum. If all binaries evolve under the influence of GW back-reaction only, in circular orbits, then the spectral index is fixed at $h_c(f) \propto f^{-2/3}$ and the GW background is fully determined (Phinney, 2001). However, to get to the point at which GW emission is efficient, MBHBs need to exchange energy and angular momentum with their stellar and/or gaseous environment (Sesana, 2013a), a process that can lead to an increase in the binary eccentricity (*e.g.* Quinlan, 1996; Cuadra et al., 2009). We assume MBHBs evolve via three-body scattering against the dense stellar background up to a transition frequency f_t at which GW emission takes over. According to recent studies (Sesana and Khan, 2015; Vasiliev et al., 2015), the hardening is dictated by the density of background stars ρ_i at the influence radius of the binary r_i. The bulge stellar density is assumed to follow a Hernquist density profile (Hernquist, 1990) with total mass M_* and scale radius a determined by the MBHB total mass $M = M_1 + M_2$ via empirical relations from the literature (see full details in Chen et al., 2017c). Therefore, for each individual system,

ρ_i is determined solely by M. In the stellar hardening phase, the binary is assumed to hold constant eccentricity e_t up to f_t, beyond which it circularises under the effect of the now-dominant GW back reaction. The GW spectrum emitted by an individual binary adiabatically inspiralling under these assumptions behaves as $h_c(f) \propto f$ for $f \ll f_t$ and settles to the standard $h_c(f) \propto f^{-2/3}$ for $f \gg f_t$. The spectrum has a turnover around f_t and its exact location depends on the binary eccentricity e_t. The observed GW spectrum is therefore uniquely determined by the binary chirp mass $\mathscr{M}$, redshift z, transition frequency f_t and eccentricity at transition e_t.

The GW spectrum from the overall population can be computed by integrating the spectrum of each individual system over the co-moving number density of merging MBHBs

$$\begin{aligned} h_c^2(f) &= \int dz \int d\log_{10}\mathscr{M} \frac{d^2n}{d\log_{10}\mathscr{M}\, dz} \\ &\quad h_{c,\mathrm{fit}}^2\left(f\frac{f_{p,0}}{f_{p,t}}\right)\left(\frac{f_{p,t}}{f_{p,0}}\right)^{-4/3}\left(\frac{\mathscr{M}}{\mathscr{M}_0}\right)^{5/3}\left(\frac{1+z}{1+z_0}\right)^{-1/3}. \end{aligned} \tag{3.5}$$

Here, $h_{c,\mathrm{fit}}$ is an analytic fit to the GW spectrum of a reference binary with chirp mass $\mathscr{M}_0$ at redshift z_0 (*i.e.* assuming $d^2n/(d\log_{10}\mathscr{M}\, dz) = \delta(\mathscr{M}-\mathscr{M}_0)\delta(z-z_0)$), characterised by an eccentricity of e_0 at a reference frequency f_0. For these reference values, the peak frequency of the spectrum $f_{p,0}$ is computed. The contribution of a MBHB with generic chirp mass, emission redshift, transition frequency f_t and initial eccentricity e_t are then simply computed by calculating the spectrum at a rescaled frequency $f(f_{p,0}/f_{p,t})$ and by shifting it with frequency mass and redshift as indicated in equation 3.5. Chen et al. (2017c) demonstrated that this simple self-similar computation of the GW spectrum is sufficient to describe the expected GW signal from a population of eccentric MBHBs driven by three-body scattering at $f > 1$ nHz, relevant to PTA measurement.

As stated above, the shape of the spectrum depends on ρ_i and e_t. The stellar density

ρ_i regulates the location of f_t; the denser the environment, the higher the transition frequency. MBHBs evolving in extremely dense environments will therefore show a turnover in the GW spectrum at higher frequency. The effect of e_t is twofold. On the one hand, eccentric binaries emit GWs more efficiently at a given orbital frequency, thus decoupling at lower f_t with respect to circular ones. On the other hand, eccentricity redistributes the emitted GW power at higher frequencies, thus pushing the spectral turnover to high frequencies. In our default model, ρ_i is fixed by the MBHB total mass M. We make the simplifying assumption that all systems have the same e_t, which although unrealistic provides a simipfied case for this initial study where the addition of parameters describing a distribution of eccentricities might provide further improvement to the model in the future.. We also consider an extended model where ρ_i is multiplied by a free parameter η. This corresponds to a simple rescaling of the central stellar density, relaxing the strict $M-\rho_i$ relation imposed by our default model. We stress here that including this parameter in our main analysis yielded quantitatively identical results.

We use a generic simple model for the cosmic merger rate density of MBHBs based on an overall amplitude and two power law distributions with exponential cut-offs,

$$\frac{\mathrm{d}^2 n}{\mathrm{d}\log_{10}\mathscr{M}\,\mathrm{d}z} = \dot{n}_0 \left(\frac{\mathscr{M}}{10^7 \mathrm{M}_\odot}\right)^{-\alpha} \exp\left(-\frac{\mathscr{M}}{\mathscr{M}_*}\right) (1+z)^{\beta} \exp\left(-\frac{z}{z_*}\right) \frac{\mathrm{d}t_r}{\mathrm{d}z} \qquad (3.6)$$

where $\mathrm{d}t_r/\mathrm{d}z$ is the relationship between time and redshift assuming a standard ΛCDM flat universe with Hubble constant of $H_0 = 70$ km s^{-1} Mpc^{-1}. The five free parameters are: $\dot{n}_0$ representing the co-moving number of mergers per Mpc^3 per Gyr; α and $\mathscr{M}_*$ control the slope and cut-off of the chirp mass distribution respectively; β and z_* regulate the equivalent properties of the redshift distribution. Equation 3.6 is also used to compute the number of emitting systems per frequency resolution bin at $f > 10$ nHz. The small number statistics of the most massive binaries determines a steepening of the GW

spectrum at high frequencies (full details of the computation are found in Sesana et al., 2008; Chen et al., 2017c). The GW spectrum is therefore uniquely computed by a set of six (seven) parameters $\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t\ (, \eta)$.

3.5.2 Anchoring the model prior to astrophysical models

Although no sub-parsec MBHBs emitting in the PTA frequency range have been unambiguously identified to date, their cosmic merger rate can be connected to the merger rate of their host galaxies. The procedure has been extensively described in Sesana (2013a). The galaxy merger rate can be estimated directly from observations via

$$\frac{\mathrm{d}^3 n_G}{\mathrm{d}z \mathrm{d}M_G \mathrm{d}q} = \frac{\phi(M_G, z)}{M_G \ln 10} \frac{F(z, M_G, q)}{\tau(z, M_G, q)} \frac{\mathrm{d}t_r}{\mathrm{d}z}. \tag{3.7}$$

Here, M_G is the galaxy mass; $\phi(M_G, z) = (\mathrm{d}n/\mathrm{dlog}M_G)_z$ is the galaxy mass function measured at redshift z; $F(M_G, q, z) = (\mathrm{d}f_p/\mathrm{d}q)_{M_G, z}$, for every M_G and z, denotes the fraction of galaxies paired with a companion galaxy with mass ratio between q and $q + \delta q$; $\tau(z, M_G, q)$ is the merger timescale of the pair as a function of the relevant parameters. We construct a library of galaxy merger rates by combining four measurements of the galaxy mass function $\phi(M_G, z)$ (Ilbert et al., 2013; Muzzin et al., 2013; Tomczak et al., 2014; Bernardi et al., 2016), four estimates of the close pair fraction $F(M_G, q, z)$ (Bundy et al., 2009; de Ravel et al., 2009; López-Sanjuan et al., 2012; Xu et al., 2012) and two estimates of the merger timescale $\tau(z, M_G, q)$ (Kitzbichler and White, 2008; Lotz et al., 2010). For each of the galaxy mass functions and pair fractions we consider three estimates given by the best fit and the two boundaries of the 1σ confidence interval reported by the authors. We therefore have $12 \times 12 \times 2 = 288$ galaxy merger rates. Each merging galaxy pair is assigned MBHs with masses drawn from 14 different MBH-galaxy relations found in the literature (for more details see table 3.4). MBHBs

are assumed to merge in coincidence with the host galaxies (*i.e.* no stalling or extra delays), but can accrete either before or after merger according to the three different prescriptions described in Sesana et al. (2009). This give a total of $14 \times 3 = 42$ distinctive MBH populations for a given galaxy merger model. We combine the 288 galaxy merger rates as per equation 3.7 and the 42 MBH masses assigned using table 3.4 plus accretion prescription into a grand total of 12096 MBHB population models. Given the uncertainties, biases, selection effects, and poor understanding on the underlying physics affecting each of the individual ingredients, we do not attempt a ranking of the models, and give each of them equal weight. The models result in an allowed MBHB merger rate density as a function of chirp mass and redshift.

We then marginalise over mass and redshift separately to obtain the functions $\mathrm{d}n/\mathrm{d}z$ and $\mathrm{d}n/\mathrm{d}\mathscr{M}$. We are particularly interested here in testing different MBH-host galaxy relations; we therefore construct the function $\mathrm{d}n/\mathrm{d}z$ and $\mathrm{d}n/\mathrm{d}\mathscr{M}$ under four different assumptions:

(i.) Model KH13 is constructed by considering both the $M-\sigma$ and $M-M_*$ relations from Kormendy and Ho (2013);

(ii.) Model G09 is based on the $M-\sigma$ relation of Gültekin et al. (2009);

(iii.) Model S16 employs the $M-\sigma$ relation from Shankar et al. (2016);

(iv.) Model ALL is the combination of all 14 of the MBH mass-host galaxy relations listed in table 3.4.

For each of these four models, the allowed regions of $\mathrm{d}n/\mathrm{d}z$ and $\mathrm{d}n/\mathrm{d}\mathscr{M}$ are shown in figure 3.5. The figure highlights the large uncertainty in the determination of the MBHB merger rate and unveils the trend of the chosen models; S16 and KH13 represent the lower and upper bound to the rate, whereas G09 sits in the middle and is representative

of the median value of model 'ALL'. These prior bands need then to be described analytically using the parameters of equation 3.6. The shape of these priors and how they differ (or not) from model to model are shown by figure 3.8.

We then ensured that once the bands of figure 3.5 are imposed on our model parameters ($\theta = \{\dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t\ (,\eta)\}$), that the resulting distribution of characteristic amplitudes h_c is consistent with that of the original models. We computed the GW background under the assumption of circular GW driven systems (*i.e.* $h_c \propto f^{-2/3}$) and compared the distributions of $h_{1\mathrm{yr}}$, *i.e.* the strain amplitudes at $f = 1\mathrm{yr}^{-1}$. The $h_{1\mathrm{yr}}$ distributions obtained with the two techniques were found to follow each other quite closely with a difference of median values and 90% confidence regions smaller than 0.1 dex. We conclude that our analytical models provide an adequate description of the observationally inferred MBHB merger rate, and can therefore be used to constrain the properties of the cosmic MBHB population. In particular model KH13 provides an optimistic prediction of the GW background with median amplitude at $f = 1\ \mathrm{yr}^{-1}$ of $h_{1\mathrm{yr}} \approx 1.5 \times 10^{-15}$; model G09 results in a more conservative prediction $h_{1\mathrm{yr}} \approx 7 \times 10^{-16}$; model S16 result in an ultra conservative estimate with median $h_{1\mathrm{yr}} \approx 4 \times 10^{-16}$; and finally the characteristic amplitude predicted by the compilation of all models (ALL) encompasses almost two orders of magnitudes with median value $h_{1\mathrm{yr}} \approx 8 \times 10^{-16}$.

As for the parameters defining the binary dynamics, we assume that all binaries have the same eccentricity for which we pick a flat prior in the range $10^{-6} < e_t < 0.999$ (see figure 3.8). In the extended model, featuring a rescaling of the density ρ_i regulating the binary hardening in the stellar phase, we assume a log flat prior for the multiplicative factor η in the range $0.01 < \eta < 100$. For detailed results of including this additional density parameter see table 3.3 and figure 3.9.

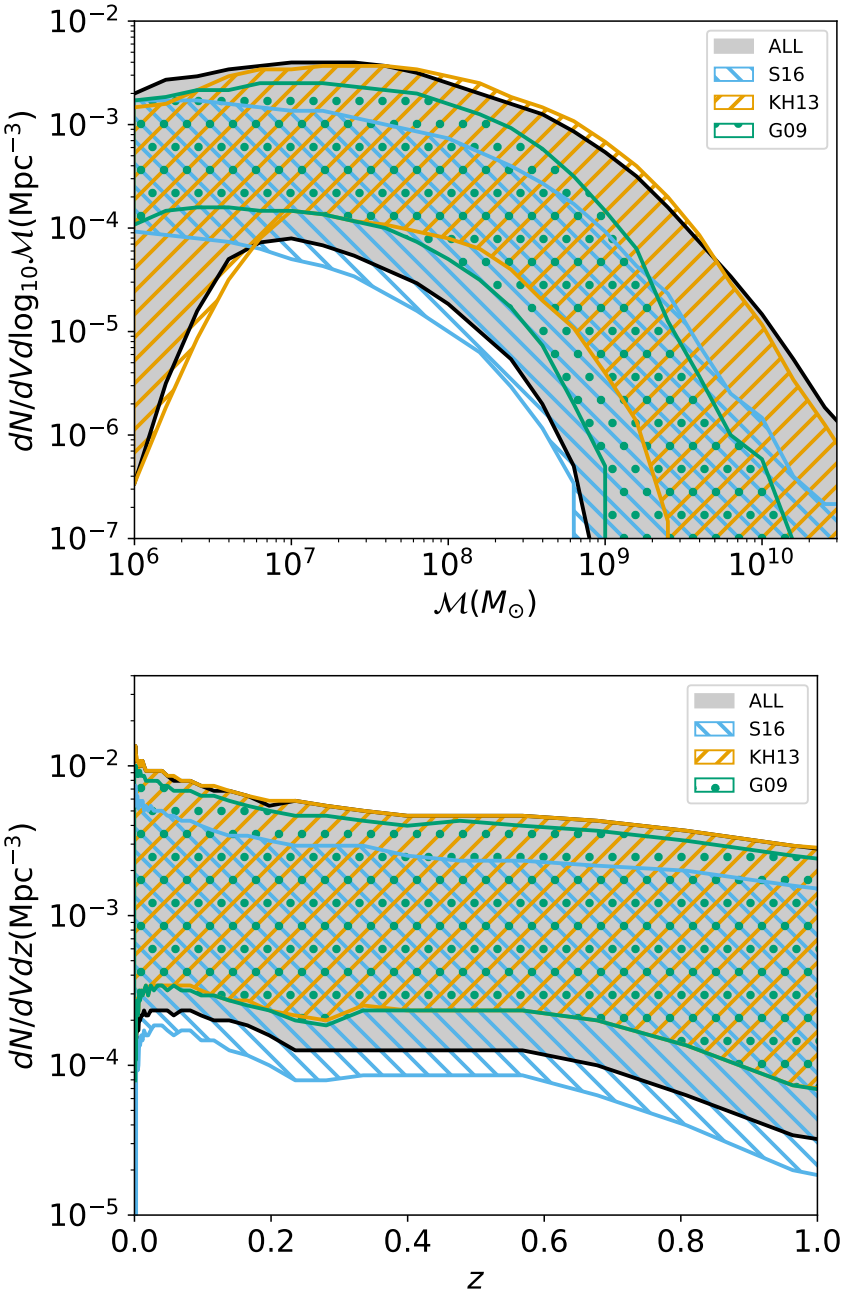


Figure 3.5: Top panel: mass density distribution $\mathrm{d}n/\mathrm{d}\mathscr{M}$ of the four astrophysical priors selected in this study (see text for full description). Bottom panel: redshift evolution of the MBHB mass density for the same four models. Note that the coloured regions represent the 99% interval allowed by each model, this is why individual models can extend beyond the region associated with model ALL (which includes KH13, G09, S16 as subsets). Reproduced from Middleton et al. (2017).

3.5.3 Likelihood function and hierarchical modelling

By making use of Bayes theorem, the posterior probability distribution $p(\theta|d,M)$ of the model parameters θ inferred by the data d given our model M is

$$p(\theta|d,M) = \frac{p(d|\theta,M)p(\theta|M)}{\mathcal{Z}_M}, \tag{3.8}$$

where $p(\theta|M)$ is the prior knowledge of the model parameters, $p(d|\theta,M)$ is the likelihood of the data d given the parameters θ and $\mathcal{Z}_M$ is the evidence of model M, computed as

$$\mathcal{Z}_M = \int p(d|\theta,M)p(\theta|M)d\theta. \tag{3.9}$$

The evidence is the integral of the likelihood function over the multi-dimensional space defined by the model parameters θ, weighted by the multivariate prior probability distribution of the parameters. When comparing two competitive models A and B, the odds ratio is computed as

$$\mathcal{O}_{A,B} = \frac{\mathcal{Z}_A}{\mathcal{Z}_B}\frac{P_A}{P_B} = \mathcal{B}_{A,B}\frac{P_A}{P_B}, \tag{3.10}$$

where $\mathcal{B}_{A,B} = \mathcal{Z}_A/\mathcal{Z}_B$ is the Bayes factor and P_M is the prior probability assigned to model M. When comparing the four models KH13, G09, S16 and ALL, we assign equal prior probability to each model. Therefore, in each model pair comparison, the odds ratio reduces to the Bayes factor. In section 3.5.2 we have defined the distribution of prior parameters $p(\theta|M)$, to proceed with model comparison and parameter estimation we need to define the likelihood function $p(d|\theta,M)$.

The likelihood function, $p(d|\theta,M)$ is defined following Chen et al. (2017a). We take the posterior samples from the Parkes PTA analysis (courtesy of Shannon and collaborators) used to place the 95% upper limit at $h_{1\mathrm{yr}} = 1\times10^{-15}$, when a single power law background $h_c \propto f^{-2/3}$ is assumed. However, for our analysis we would like

to convert this upper limit at $f = 1\text{yr}^{-1}$ to a frequency dependent upper limit on the spectrum as shown by the orange curve in figure 3.3. Our likelihood is constructed by multiplying all bins together, therefore the resulting overall limit from these bin-by-bin upper limits must be consistent with $h_{1\text{yr}} = 1 \times 10^{-15}$. The $f_{1\text{yr}}$ posterior distribution is well fitted by a Fermi function. To estimate a frequency dependent upper limit, we use Fermi function likelihoods at each frequency bin, which are then shifted and re-normalised in order to provide the correct overall upper limit. In our analysis we consider the contributions by only the first four frequency bins of size $1/11$ yr^{-1}, as the higher frequency portion of the spectrum provides no additional constraint. We have verified that when we include additional bins the results of the analysis are unchanged. Ideally, we would take the bin-by-bin upper limits directly from the pulsar timing analysis to take account of the true shape of the posterior; however, the method we use here provides a consistent estimate for our analysis.

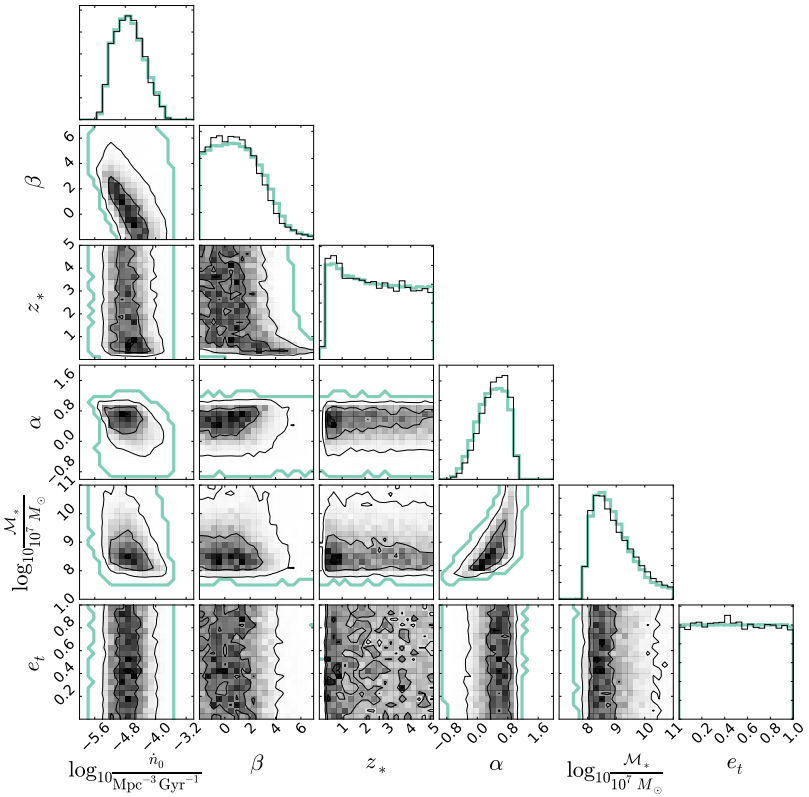
Having defined the population of merging binaries, the astrophysical prior and the likelihood based on the PPTA upper limit result, we use a nested sampling algorithm (Skilling, 2004; Veitch et al., 2017) to construct posterior distributions for each of the six model parameters. For the results shown here, we use 2000 live points and run each analysis 5 times, giving an average of around 18000 posterior samples.

3.6 Supplementary Material

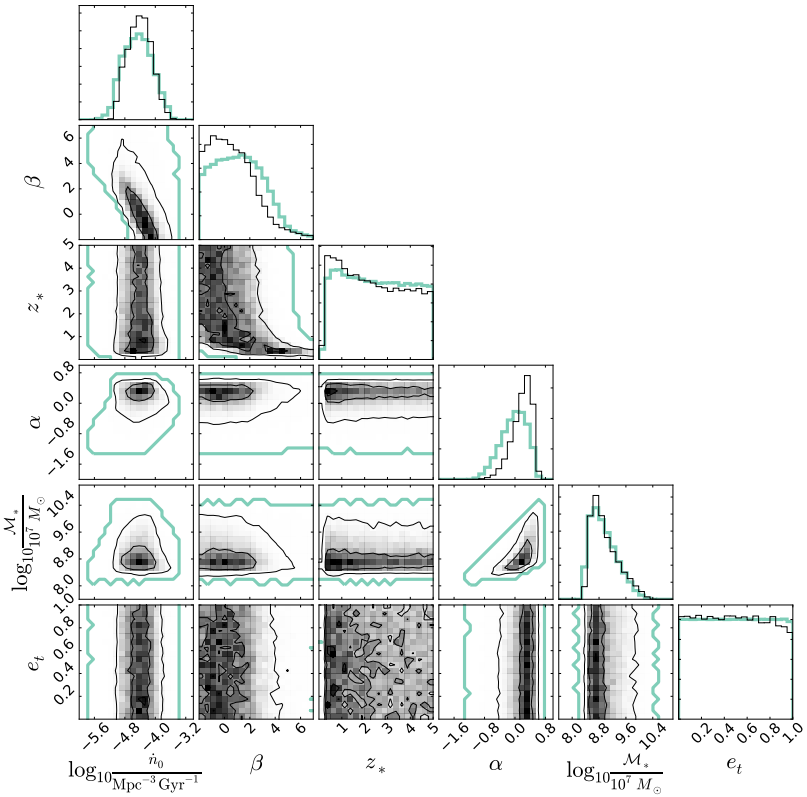
The following pages show supplementary material which support the paper Middleton et al. (2017). The figures, tables and text are taken directly from the paper, but may still be edited before publication.

Figure 3.6: Plots are displayed on following two pages. The nested sampling algorithm (section 2.2.2) returns the full posteriors of the N-dimensional parameter space and the model evidences (see table 3.3). Here we show the posteriors for our main analysis of the PPTA upper limit using the default six parameter model ($\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t$). The four panels show the result for each of the astrophysical model priors: S16; KH13; G09; ALL. Within each panel, the plots on the diagonal show the one-dimensional marginalised distributions for each parameter with the posterior (thin-black) and prior (thick-green). The central plots show the two-dimensional posterior distributions for each parameter pair (grey-shading) along with the extent of the prior (thick-green contour).
It is clear by comparison of the prior and posterior that the current PTA observations impose little constraint on the shape of the MBHB mass function. For S16, the most conservative model, the prior and posterior are virtually identical. Even for the KH13 model, the two distributions match closely, with only appreciable differences for β and α due to the PPTA limit excluding the highest values of h_c predicted by the model (see also figure 3.3), resulting in a preference for large α and negative β. In fact, for the mass function adopted in equation 3.6, a large α results in a MBHB population dominated by low mass systems, which tends to suppress the signal. Likewise, a small (or negative) β implies a sparser population of MBHB at higher redshift, again reducing the GW background level. In any case, current PTA measurements provide little new information on the MBHB cosmic population as also demonstrated by the small K-L divergences between prior and posterior of individual model parameters in table 3.2. Reproduced from the supplementary material of Middleton et al. (2017).

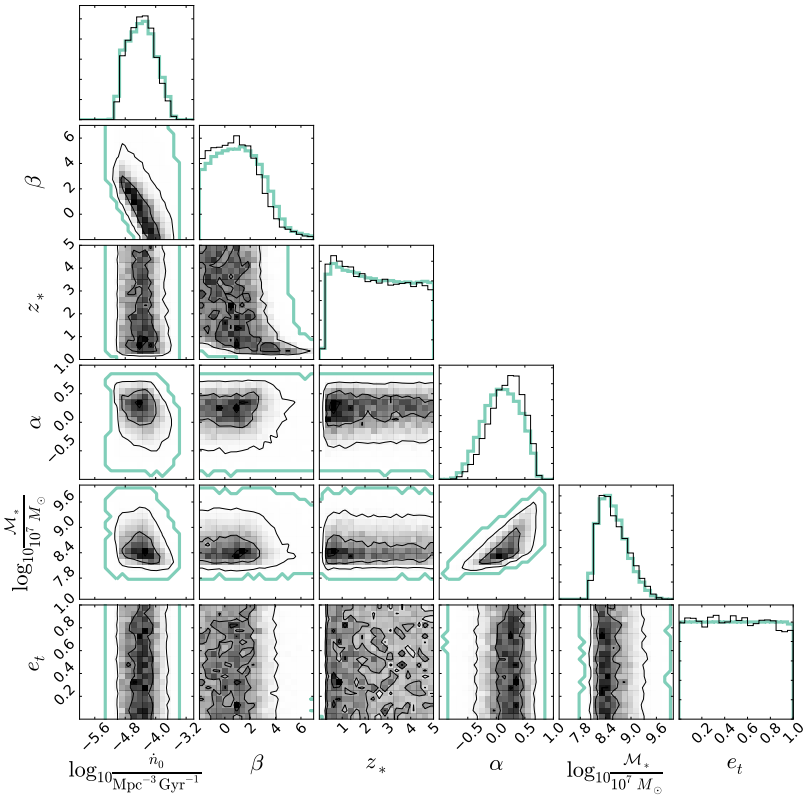
S16


$\log_{10}\frac{\dot{n}_0}{\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}}$
β
z_*
α
$\log_{10}\frac{\mathcal{M}_*}{10^7\,M_\odot}$
e_t

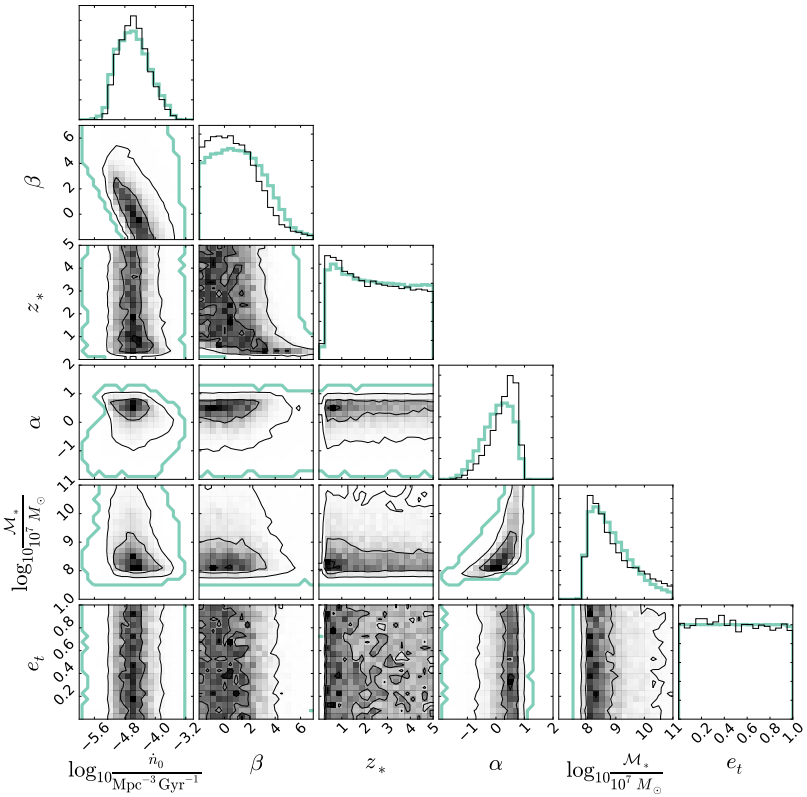
KH13


$\log_{10}\frac{\dot{n}_0}{\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}}$
β
z_*
α
$\log_{10}\frac{\mathcal{M}_*}{10^7\,M_\odot}$
e_t

G09

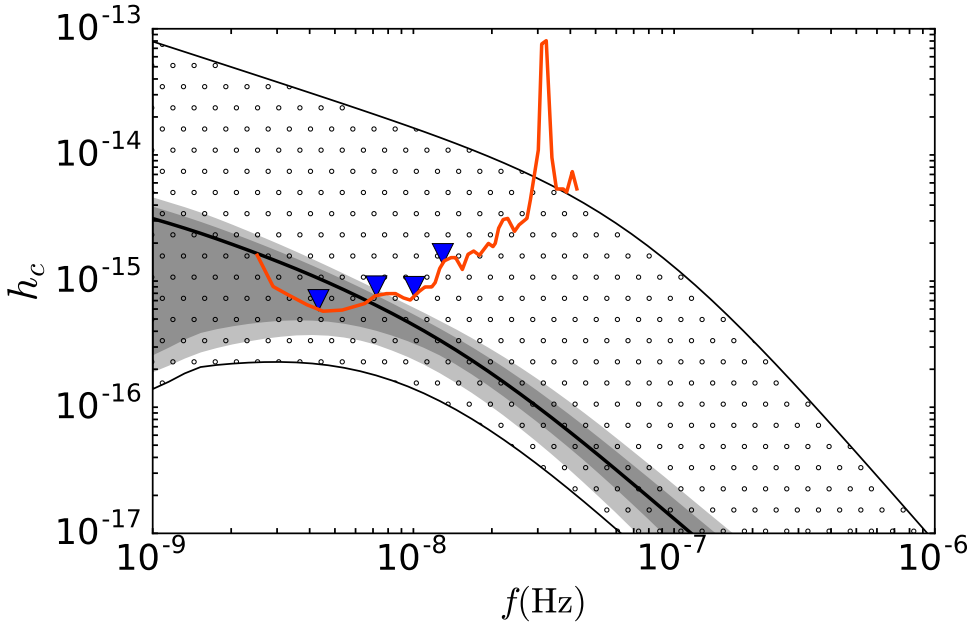

$\log_{10}\frac{\dot{n}_0}{\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}}$
β
z_*
α
$\log_{10}\frac{\mathcal{M}_*}{10^7\,M_\odot}$
e_t

ALL

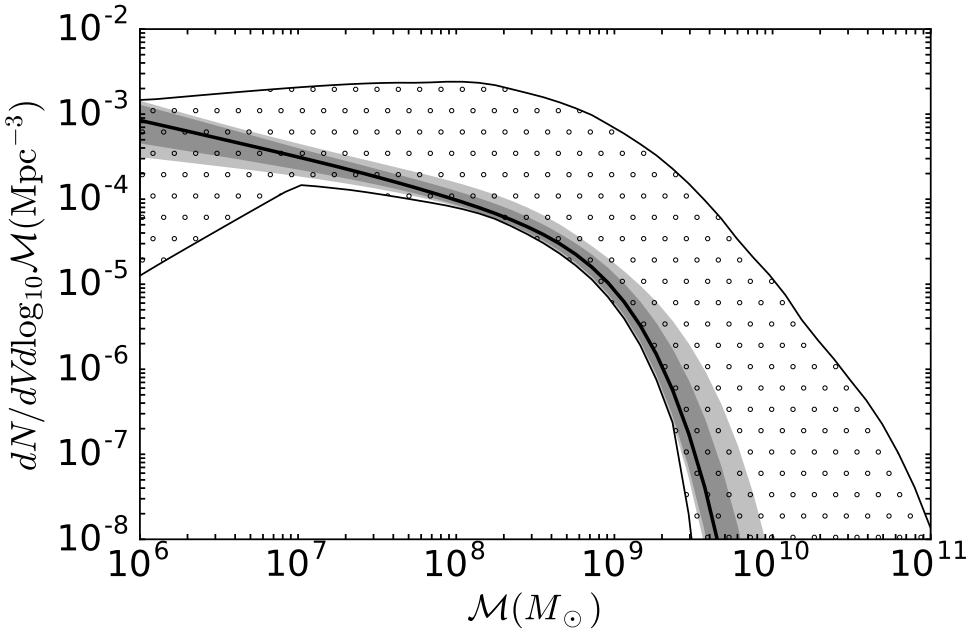

$\log_{10}\frac{\dot{n}_0}{\mathrm{Mpc}^{-3}\,\mathrm{Gyr}^{-1}}$
β
z_*
α
$\log_{10}\frac{\mathcal{M}_*}{10^7\,M_\odot}$
e_t

Model	K-L divergence					
	$\log_{10}\dot{n}_0$	β	z_*	α	$\log_{10}\mathscr{M}_*$	e_t
KH13	0.06	0.05	< 0.01	0.24	0.03	< 0.01
G09	< 0.01	0.01	< 0.01	0.04	0.01	< 0.01
S16	< 0.01	< 0.01	< 0.01	0.01	< 0.01	< 0.01
ALL	0.02	0.02	< 0.01	0.08	0.02	< 0.01

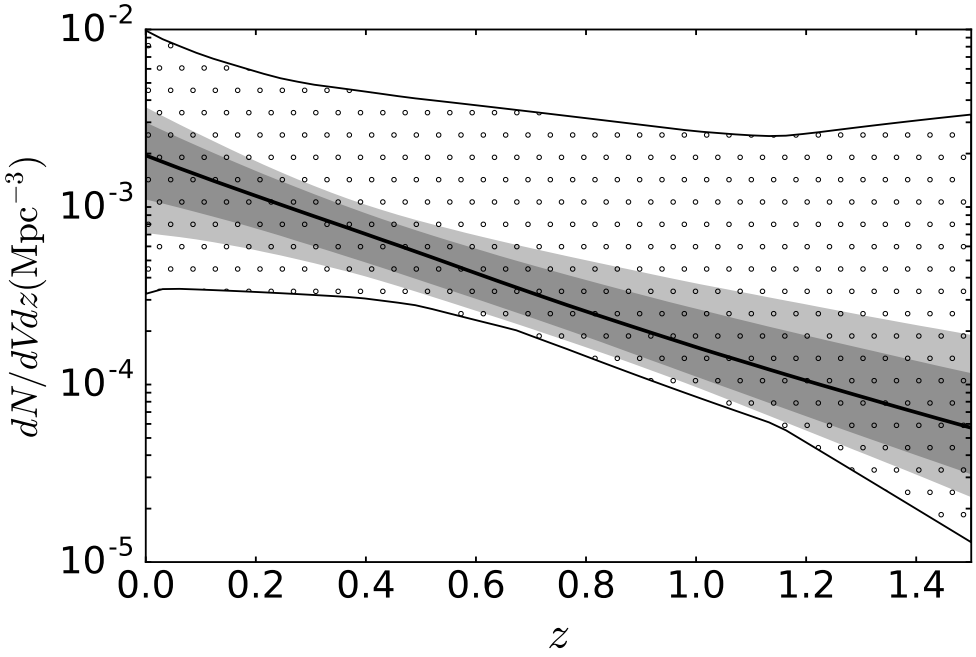
Table 3.2: The K-L divergences in the table compare the marginalised prior and posterior distributions for the individual parameters in the default 6-parameter model ($\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t$) and given the PPTA upper limit. The K-L divergence is a measure of the degree of similarity between two probability distributions with zero being identical. The rows indicate the four astrophysical models used as priors in this study (KH13, G09, S16, ALL). As also seen qualitatively from figure 3.6, there is little difference between the prior and posterior for all four models. KH13, the model predicting the highest characteristic strain as expected has the highest K-L divergences, however as also shown by figure 3.6 the difference is still small. Reproduced from the supplementary material of Middleton et al. (2017).



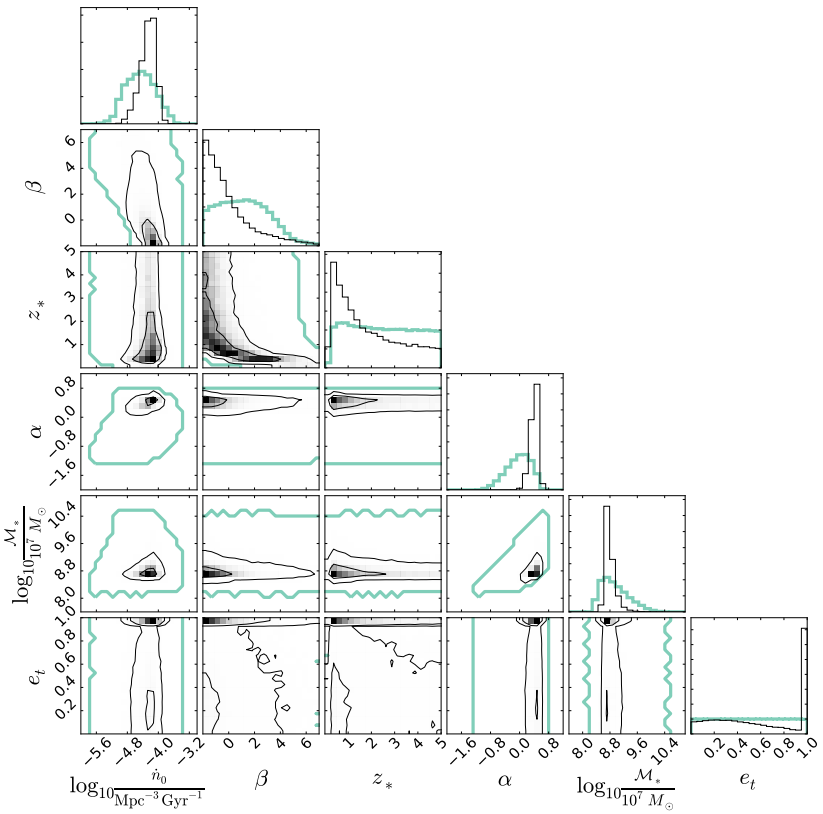
(a) The characteristic amplitude prior (dotted band) and posterior (shaded bands) The shading represents the 68% and 90% credible bands of the posterior with the solid line showing the median value. The orange curve shows the bin-by-bin 95% upper limits at different frequency bins (with blue triangles indicating the frequency bins we use), resulting in an overall limit $h_{1\mathrm{yr},95\%} = 1 \times 10^{-16}$. See also the main caption at end of these subfigures.



(b) Here we show the prior and posterior distributions for the mass density function. As in panel 3.7a, we represent the prior with the dotted area and the shaded bands show the 68% and 90% credible bands of the posterior, with the median marked as the solid line. See also the main caption at end of these subfigures.



(c) Identical to subplot 3.7b above, but for the redshift evolution of the MBHB population. See also the main caption at the end of these subfigures.



(d) The individual posterior density functions for each of the model parameters. The diagonal plots show the one-dimensional posterior distribution (black) along with the prior (green-thick), whilst the central plots show the two-dimensional posterior for each of the parameter pairs again with the extent of the prior shown by the single green-thick contour. See also the main caption at the end of these subfigures.

Figure 3.7: Subfigures are also displayed on the previous two pages. Here we show results of a upper limit of $h_{1\mathrm{yr},95\%} = 1\times10^{-16}$ for the six parameter model ($\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t$) and astrophysical prior KH13. We see here that now all the posteriors differ significantly from the respective prior. From panel 3.7d, low β and z_* are preferred, because this suppresses the total number of MBHs at high redshifts. Higher values of $\dot{n}_0$ are preferred. Although this might be surprising, it is dictated by the shape of the prior of $\mathrm{d}n/\mathrm{d}z$ (panel 3.7c); in order to minimise the signal, it is more convenient to allow a negative β at the expenses of a higher local normalisation $\dot{n}_0$ of the merger rate. High α values are preferred, since they imply a population dominated by low mass MBHBs (panel 3.7b). The e_t posterior now shows a prominent peak close to the maximum $e_t = 0.999$, with a long tail extending to zero (right-most plot of panel 3.7d). Very high eccentricities are preferred, although low values are still possible. This is because 1×10^{-16} is only a 95% upper limit, therefore there is a small chance that a low eccentricity model producing a signal surpassing the 1×10^{-16} value is nonetheless accepted in the posterior. Reproduced from the supplementary material of Middleton et al. (2017).

Model	$h_{1\mathrm{yr},95\%} = 1\times10^{-15}$		$h_{1\mathrm{yr},95\%} = 3\times10^{-16}$		$h_{1\mathrm{yr},95\%} = 1\times10^{-16}$	
	e_t	$e_t+\eta$	e_t	$e_t+\eta$	e_t	$e_t+\eta$
KH13	$-2.36\,(0.85)$	$-2.23\,(0.84)$	$-5.68\,(2.25)$	$-5.47\,(2.25)$	$-13.17\,(5.18)$	$-9.03\,(7.11)$
G09	$-1.2\,(0.39)$	$-1.1\,(0.39)$	$-3.35\,(1.11)$	$-3.17\,(1.09)$	$-8.26\,(2.86)$	$-6.38\,(4.02)$
S16	$-0.6\,(0.37)$	$-0.57\,(0.38)$	$-1.62\,(0.69)$	$-1.6\,(0.71)$	$-3.82\,(1.42)$	$-3.56\,(1.48)$
ALL	$-1.23\,(0.62)$	$-1.14\,(0.62)$	$-2.68\,(1.33)$	$-2.63\,(1.31)$	$-5.74\,(2.50)$	$-5.09\,(2.53)$

Table 3.3: The table shows the natural logarithm of model evidences and associated h_c K-L divergences (in parenthesis) for each of the astrophysical models: KH13, G09, S16 and ALL. We consider two different parametrisations of the MBHB dynamics; one which has only e_t as a free parameter (column 'e_t', the 6 parameter model), and one where we add the normalisation factor η to the density at the influence radius ρ_i as a free parameter (column '$e_t+\eta$', the 7 parameter model). Numbers are reported for three values of the 95% PTA upper limit; the current upper limit at 1×10^{-15} and two possible future upper limits at 3×10^{-16} and 1×10^{-16}. See also the Note below. Reproduced from the supplementary material of Middleton et al. (2017).

Note: Progression of constraints with future possible upper limits

From table 3.3 we see that, given the current upper limit of $h_{1\mathrm{yr},95\%} = 1\times10^{-15}$, there are no significant differences between the 6 and 7 parameter results, with virtually identical evidences and K-L divergences. Together with the flat e_t posteriors shown in figure 3.6, this leads us to an important conclusion: current PTA non detections do not favour (nor require) a strong coupling with the environment. Neither high stellar densities (*i.e.* efficient 3-body scattering) nor high eccentricities are preferred by the data. The conservative S16 model is always favoured, however even when compared to KH13, one obtains $\ln\mathcal{B} = 1.76$, which only mildly favours S16 (Kass and Raftery, 1995). In addition, all K-L divergences are smaller than unity, indicating only minor updates with respect to the h_c prior distributions. This is another measure of the fact that the data are not very informative.

Turning to the limit at $h_{1\mathrm{yr},95\%} = 3\times10^{-16}$, the K-L divergences of all models,

with the exception of S16, are now larger than unity indicating that the upper limit is becoming more informative. In terms of model comparison, S16 is now mildly favoured with respect to G09 ($\ln \mathcal{B} = 1.73$) and strongly favoured compared to KH13 ($\ln \mathcal{B} = 4.06$). Again, we note that adding η does not make a significant difference to the model evidence. Even with such a low upper limit, neither high eccentricity nor strong coupling with the environment improve the agreement between model expectations and data. Although this seems counter-intuitive, we should keep in mind that the upper limit is set around $f \approx 5 \times 10^{-9}$ Hz (see figure 3.3). Any dynamical effect should therefore cause a turnover of the spectrum around 10^{-8} Hz to have an impact on model selection, which occurs only in a small corner of parameter space where both e_t and η are high. However, for all models $h_{1\mathrm{yr},95\%} = 3 \times 10^{-16}$ is still consistent with the tail of the h_c distribution when an $f^{-2/3}$ spectrum is assumed. Invoking high e_t and η is not necessary.

The limit becomes far more interesting if it reaches $h_{1\mathrm{yr},95\%} = 1 \times 10^{-16}$. Now all K-L divergences are substantial, indicating that the measurement is indeed informative. Model selection now strongly favours model S16 compared to any other model, whether η is included or not. Even including all environmental effects, we find that S16 is decisively preferred over KH13 with $\ln \mathcal{B} = 5.47$. Note however, that S16 has a log evidence of -3.56 of its own, considerably lower than zero (the evidence of a model is unaffected by the measurement). Since delays and stalling can potentially decrease the GW background by preventing many MBHB from merging, it is likely that a non detection at this level will provide strong support for those dynamical effects. These are not yet included in our modelling and we plan to explore them in the future.

Note reproduced from the supplementary material of Middleton et al. (2017).

Paper	X	a	b	ε
Häring and Rix (2004)	M_*	8.2	1.12	0.30
Sani et al. (2011)	M_*	8.2	0.79	0.37
Beifiori et al. (2012)	M_*	7.84	0.91	0.46
McConnell and Ma (2013)	M_*	8.46	1.05	0.34
Graham (2012)	M_*	8.56	1.01	0.44
		(8.69)	(1.98)	(0.57)
Kormendy and Ho (2013)	M_*	8.69	1.17	0.29
Sani et al. (2011)	L_i	8.19	0.93	0.38
Gültekin et al. (2009)	σ	8.23	3.96	0.31
Graham et al. (2011)	σ	8.13	5.13	0.32
Beifiori et al. (2012)	σ	7.99	4.42	0.33
McConnell and Ma (2013)	σ	8.33	5.57	0.40
Graham and Scott (2012)	σ	8.28	6.01	0.41
Kormendy and Ho (2013)	σ	8.5	4.42	0.28
Shankar et al. (2016)	σ	7.8	4.3	0.3

Table 3.4: For each merging galaxy pair, we assign a black holes with masses drawn from 14 MBH-galaxy relations found in the literature. In the table we list the relations used. The total black hole mass is described by the relation $\log_{10} M = a + b\ \log_{10} X$. Here $X = \{\sigma/200\text{km s}^{-1}, L_i/10^{11} L_\odot \text{ or } M_*/10^{11}\ \text{M}_\odot\}$, with σ being the stellar velocity dispersion of the galaxy bulge, L_i its mid-infra-red luminosity, and M_* its bulge stellar mass. Each relation is also characterised by an intrinsic scatter ε. In the table, we list the parameters a, b and ε for each of the relations taken from the literature. There are two entries for Graham (2012) who proposes a double power law with a break at $\bar{M}_* = 7 \times 10\ \text{M}_\odot$, for which the values in parenthesis refer to $M_* < \bar{M}_*$. Reproduced from the supplementary material of Middleton et al. (2017).

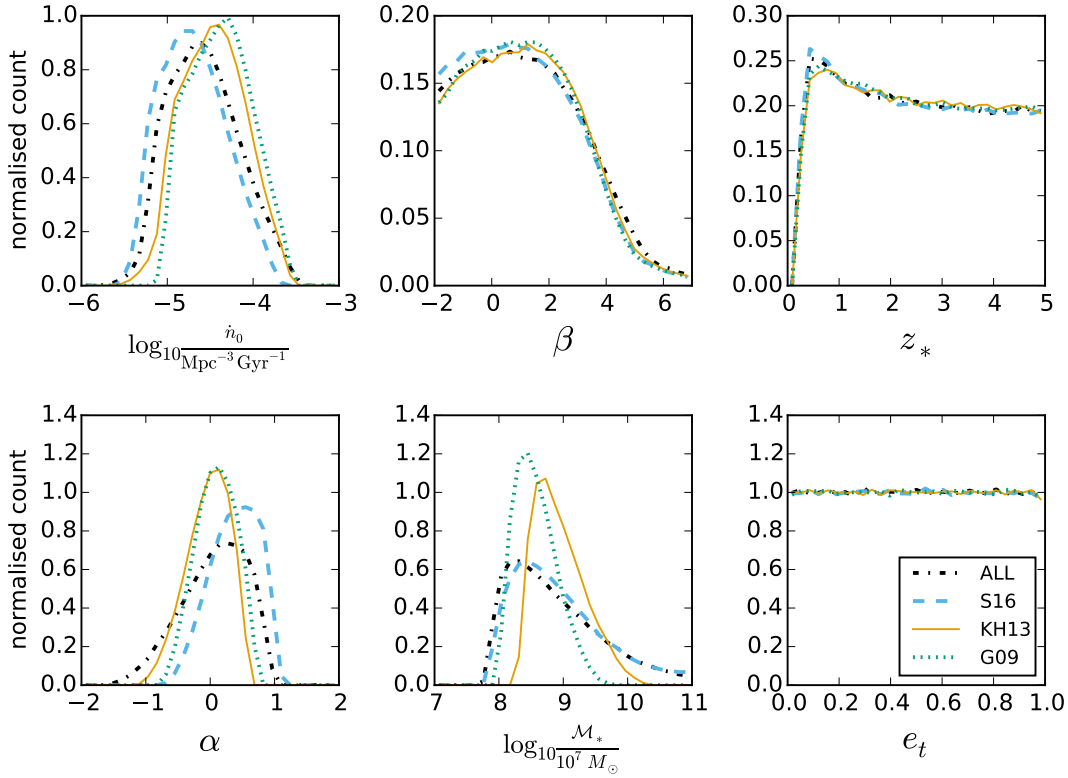
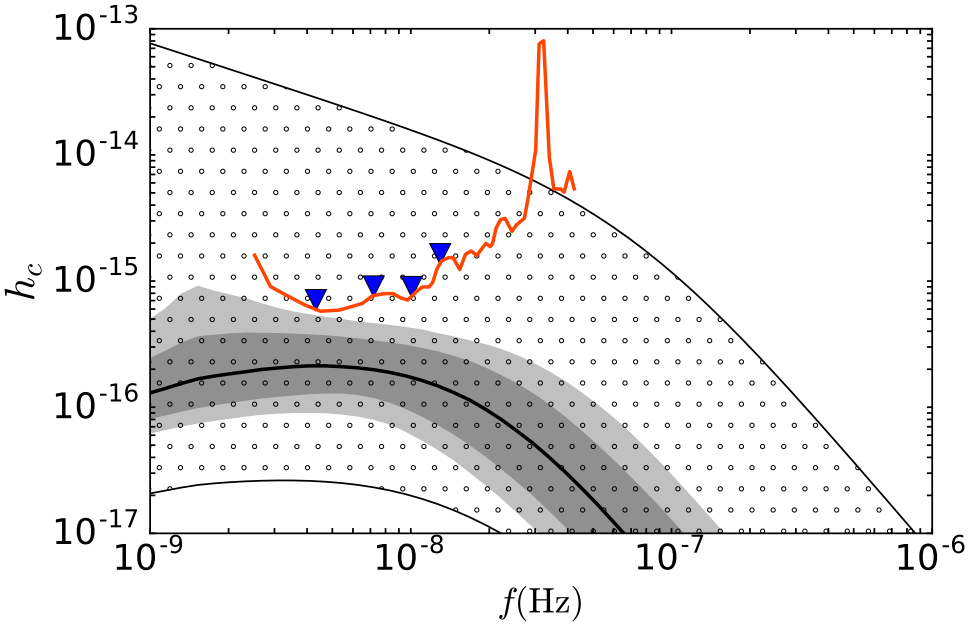
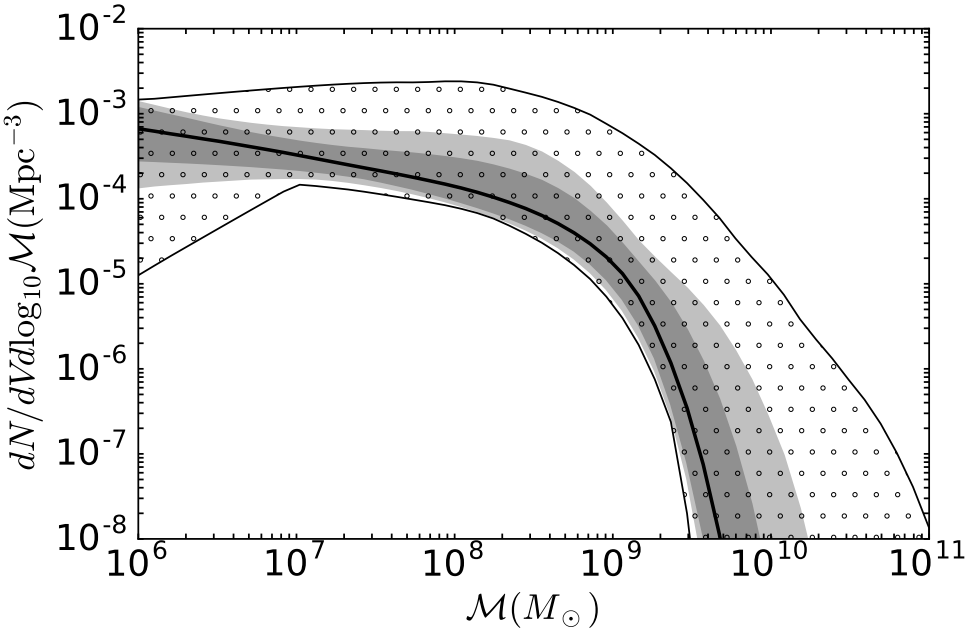


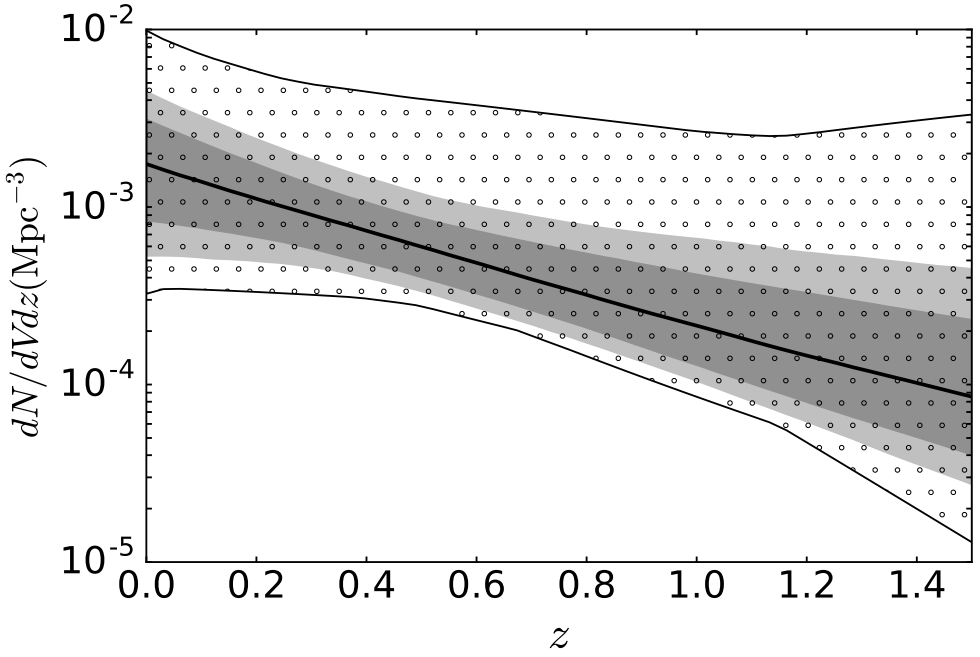
Figure 3.8: Here we show the prior density functions for each of the individual model parameters and for each astrophysical model. The numerical MBHB mass functions obtained from the prior bands in figure 3.5 need to be related analytically to the parameters of equation 3.6. Our strategy is therefore to make a large series of random draws of the five parameters defining equation 3.6, and to retain only those sets that produce $\mathrm{d}n/\mathrm{d}z$ and $\mathrm{d}n/\mathrm{d}\mathscr{M}$ within the boundaries set by the empirical models shown in figure 3.5. The prior distributions for the individual parameters obtained in this way are shown here. Top row from left to right $\dot{n}_0$, β, z_*; bottom row from left to right α, $\mathscr{M}_*$, e_t. The lines represent the priors for the four astrophysical models KH13 (orange-solid), S16 (blue-dashed), G09 (green-dotted) and ALL (black-dash-dot). We see that the redshift parameters (β and z_*) have a very similar prior for each of the models. The main differences are seen in the number rate density of mergers $\dot{n}_0$ and in the mass distribution parameters (α and $\mathscr{M}_*$). KH13 and ALL prefer higher values of $\dot{n}_0$. S16 allows for slightly higher values of α (in comparison to KH13 and G09), corresponding to a more negative slope on the mass distribution, with preference for a larger number of low mass binaries. The eccentricity parameter e_t is unaffected by the choice of model, therefore we place on it a flat prior in the range $10^{-6} < e_t < 0.999$. Reproduced from supplementary material of Middleton et al. (2017).



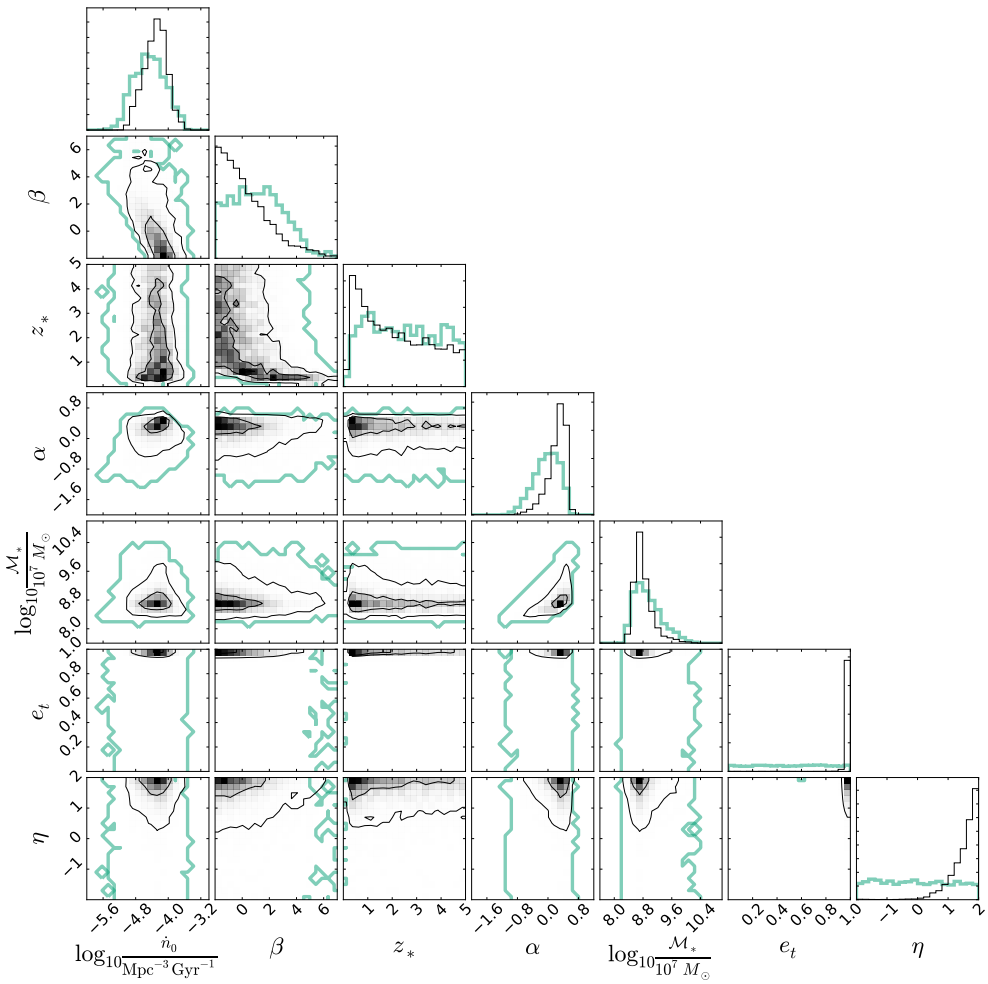
(a) Identical to panel (a) of figure 3.7 except for the addition of the stellar density parameter. See also the main caption at end of these subfigures.



(b) Identical to panel (b) of figure 3.7 except for the addition of the stellar density parameter. See also the main caption at end of these subfigures.



(c) Identical to panel (c) of figure 3.7 except for the addition of the stellar density parameter. See also the main caption at the end of these subfigures.



(d) Identical to panel (d) of figure 3.7 except for the addition of the stellar density parameter. See also the main caption at the end of these subfigures.

Figure 3.9: Subfigures are also displayed on the previous two pages. Here we show results of a upper limit of $h_{1\mathrm{yr},95\%} = 1 \times 10^{-16}$ for the seven parameter model ($\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t, \eta$) and astrophysical prior KH13. The plots shown here are identical to those of figure 3.7, the only difference being the addition of the stellar density parameter η. Most notably, we see now that extremely high eccentricities and high densities are strongly favoured. This is primarily because the addition of η extends the prior in h_c (shown in the upper left panel) downwards quite below the level imposed by the upper limit. It is therefore now easier to find points in the parameter space consistent with the measurement when e_t and η are large. Should other MBH-host galaxy relations being ruled out by independent constraints, a PTA upper limit of 1×10^{-16} would provide strong evidence of surprisingly extreme dynamical conditions of MBHBs. Reproduced from the supplementary material of Middleton et al. (2017).

Chapter 4

Tests of General Relativity

This chapter contains work in progress towards a paper on the subject of general tests of theories. The work is led by Walter Del Pozzo and done in collaboration with co-authors listed below. In this work I have contributed towards discussions, early versions of the code and ran the analysis for Newton's constant of gravitation. The results presented here are preliminary and the final published form of the paper will differ from the text reproduced here.

Authorship:
Walter Del Pozzo, Hannah Middleton, Alberto Vecchio, Jonathan Gair, Ilya Mandel

4.1 Testing theories when there are no alternatives

In Bayesian analysis, in order to make statements on the relative probability of one model compared to another, we perform model selection (see section 1.8.1). However, this relies on the existence of multiple models to weigh up against each other, *i.e.* the result is a comparison between models rather than a statement on the individual model itself. Here we consider the scenario in which there is no viable alternative model to consider – what can be said about the validity of the model when put under experimental test?

We place a particular focus on the case of general relativity (GR) here, as a prime example of a theory which has so far passed every test applied to it but is incompatible with quantum mechanics. Many individual experiments have searched for deviations from the predictions of GR, testing different sectors of the theory.

Here we describe a means to combine the results of multiple experiments testing different aspects of the same theory, with the aim of computing a 'global *p-value*' for a theory. As any violation of a theory may be at low significance (particularly in the case of GR), it can be more informative to combine results to make an overall statement on the theory – if there are multiple experiments each with some small degree of violation, their combination will produce a more significance statement on any inconsistency. The underlying statistical method we use is not new, but is based on the century-old Fishers combined probability test (Fisher, 1970).

In section 4.2 we cover the method for combining experiments; we then apply the method to a gedankan set-up in section 4.3 and to real scenarios in section 4.4. Sections 4.2, 4.3 and 4.4.1 are all reproduced in edited form from Del Pozzo et al. (in prep.).

4.2 Method

We have a theory we would like to test, which we call T. The theory makes a prediction for the value of some parameter x to be x_0. Our starting point is the result of an analysis which produces a posterior probability distribution for parameter x. The predicted value x_0 will lie within some credible interval of the posterior, which can be calculated by integration of the posterior $p(x|d,T)$

$$q = \int_{-\infty}^{x_0} p(x|d,T)dx. \tag{4.1}$$

Then, if there are N independent experiments conducted, the credible interval (a value between 0 and 1) at which the value predicted by the theory can be computed for each, giving q_i where $i = 1...N$. Switching from a credible interval to a *p-value*, we choose

$$p_i \;=\; 2\min\left(q_i, 1-q_i\right). \tag{4.2}$$

The minimum of q_i and $1-q_i$ is chosen in order to remove the effect of choosing the direction of integration (*i.e.* $-\infty \to x_0$ or $x_0 \to \infty$). If the theory T is correct, the *p-values* for N independent tests should be affected by only statistical fluctuations and should therefore be uniformly distributed as $\mathcal{U}[0,1]$.

Once we have this set of N experiments reduced to one *p-value* from each, we would like to combine them into a single value. Fisher showed if $p_i \sim \mathcal{U}[0,1]$, then the quantity (Fisher, 1970),

$$\chi^2_{2N} = -2\sum_{i=1}^{N}\ln\left(p_i\right), \tag{4.3}$$

is chi-squared distributed with $2N$ degrees of freedom, giving a means to combine experiments into a single statistic.

One advantage of this method is that the representation of a single experiment as a

single *p-value*, which is independent of the measured parameter, enables the combination of experiments testing very different aspects of the same theory. An overall statement of the theory as a whole can therefore be made.

4.3 Gedankan Experiment

Here we construct a gedankan experiment to illustrate the method. There are N independent experiments, each summarised by the quantile at which the predicted theory value is found. We consider two alternatives – either the theory T is true which we call hypothesis H_0, or T is not true, which we name as H_1. For simplicity, we assume that if H_0 is true, then the posterior distribution of the parameter in question will be normally distributed

$$P(x|d,H_0,I) \sim \mathcal{N}[\mu_0,\sigma], \tag{4.4}$$

however, if H_1 is true, then the posterior distribution will follow a different distribution

$$P(x|d,H_1,I) \sim \mathcal{N}[\mu_1,\sigma], \tag{4.5}$$

where σ, the standard deviation of the distribution reflects the precision of the experiment conducted. In this simple formalism we choose $\sigma = 1$ and the resolving power of the method is decided by the size of the deviation, or difference in the two distribution, which we call ε so that $\mu_1 = \mu_0(1+\varepsilon)$

For the scenario constructed here, if hypothesis H_0 is the correct theory then the quantiles at which x_0 is found will be uniformly distributed $\mathcal{U}[0,1]$ However, if the hypothesis H_1 is true, then by falsely assuming that x should be distributed according to the H_0 prediction as in equation 4.4, the quantiles will not be uniformly distributed according to $\mathcal{U}[0,1]$.

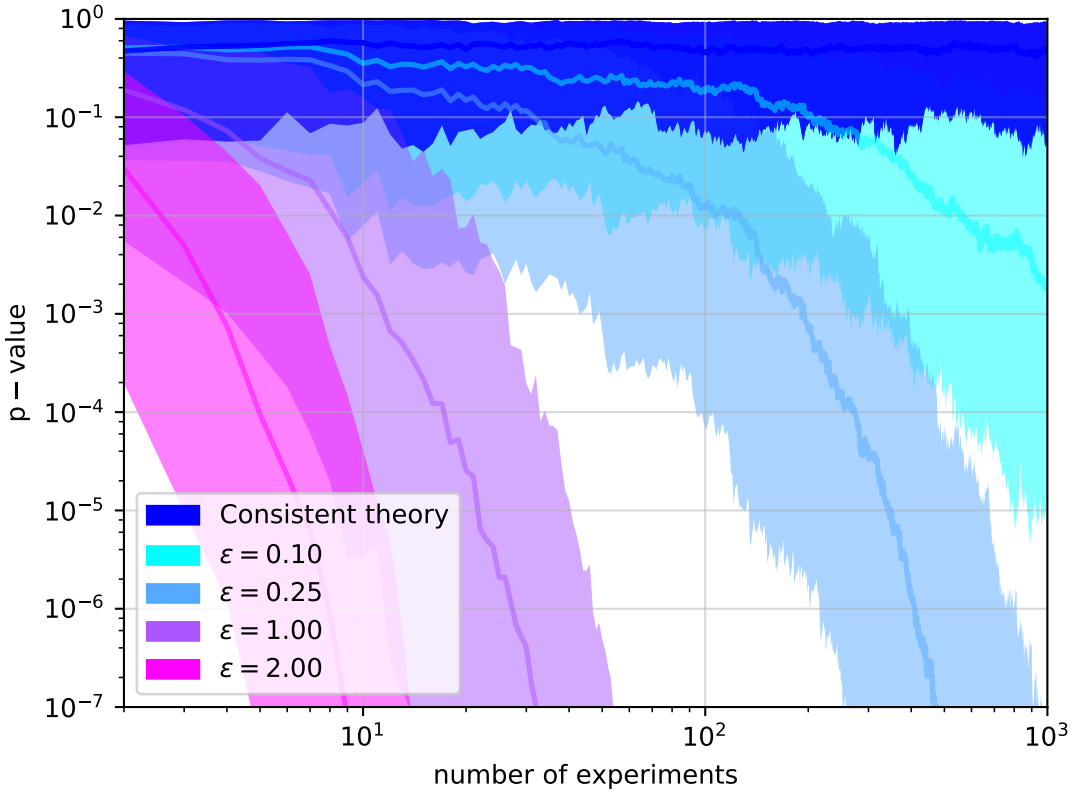


Figure 4.1: Showing the *p-values* calculated for increasing numbers of experiments. The different colour banding shows different levels of deviations (at 10%, 25%, 100%, 200%) from the theory under consideration. For each level of deviation, 1000 realisations are used to construct the median (solid line) and the 90% confidence band (coloured shading). We see that the level of the deviation is an important factor in how soon any violation becomes apparent and that for small deviations, there may be no conclusive difference even up to 1000 experiments. Reproduced from Del Pozzo et al. (in prep.).

If only one experiment is carried out it is of course impossible to provide a convincing statement on the theory as a whole with this method, however as more experiments are added to the collection, stronger statements can be made. The effect of the number of experiments on the *p-values* calculated can be seen in figure 4.1. The theory H_0 is represented by the blue band and the other coloured bands show the case where the quantiles were actually drawn from the other hypotheses with different degrees of deviation from the H_0 prediction ($\varepsilon = 10\%$, 25%, 100%, 200%). Each band is constructed from 1000 realisations. The solid lines show the median of these realisations and the extent of the bands represents the 90% credible regions. It can be seen from figure 4.1 that the larger the deviation, the sooner (in terms of number of experiments) the violation becomes clear. The figure also shows the power of combining multiple experiments. For the case of only a 1σ violation, individual results are not enough to be able to say there is any indication of a problem with the theory. However, by combining multiple experiments, this becomes more apparent after ~ 10 experiments and very clear for ~ 50 experiments. On the other hand, if the deviation is small ($\varepsilon = 10\%$) then the results are inconclusive even up to 1000 experiments (cyan band).

It might also be desirable to choose a threshold beyond which concerned might be raised over a theory, although the choice of this value is subjective. Figure 4.2 shows the number of experiments necessary to reach a certain *p-value* (the thresholds chosen are either 10^{-5} or 10^{-7}) for a given level of violation. The smaller the violation, the more experiments are needed to provide an indication of a deviation and for extremely small deviations (10^{-1}–10^{-2}), it is not possible to reach either threshold with ~ 1000 experiments, reiterating the results of figure 4.1.

In the following section, we list two examples where we have applied this method, however it is sufficiently generic that it can be used to test any theory where individual results can be reduced to a *p-value*.

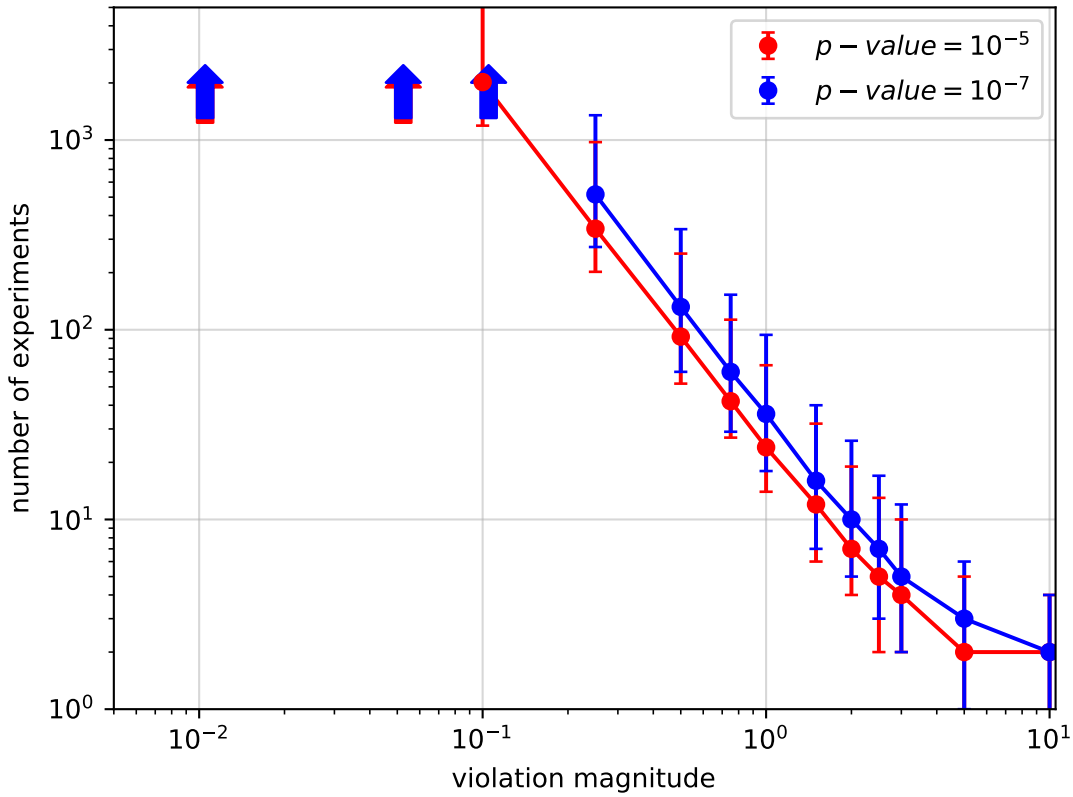


Figure 4.2: The number of experiments needed to result in a threshold *p-value* of 10^{-5} (red) and 10^{-7} (blue) for a given magnitude of violation. The error bars represent the 90% uncertainty produced from 1000 realisations. Reproduced from Del Pozzo et al. (in prep.).

4.4 Applications

Here we show two examples of the application of this method. Both are related to gravity, however the method could equally be applied to other theories under scrutiny where no alternative theory exists for comparison. In section 4.4.1 we consider the case of general relativity, combining results of the recent observations of gravitational waves from merging black hole binaries (Abbott et al., 2016j,g, 2017b) and in section 4.4.2 we apply the method to Newton's constant (G), the precise measurement of which has proved elusive when comparing many experimental results (Wood, 2014; Mohr et al., 2012).

4.4.1 Gravitational wave observations

The results of the tests of general relativity (GR) with the first gravitational wave observations can be found in Abbott et al. (2016j), Abbott et al. (2016g) and Abbott et al. (2017b). Here we provide a brief summary of their method before calculating an overall *p-value* from these results.

Parametrised tests with gravitational wave from binary black hole mergers

The post-Newtonian (PN) formalism is an approximation to Einstein's equation of gravity as an expansion in the small parameter, v/c (velocity as a fraction of the speed of light). The gravitational-wave strain h for the inspiral phase of a binary merger with chirp mass $\mathscr{M}$ (equation 1.9) and luminosity distance D_L is given by,

$$h(f) = \frac{1}{\sqrt{30}\pi^{2/3}} \frac{\mathscr{M}^{5/6}}{D_L} f^{-7/6} \exp\left[i \left(2\pi f t_c - \Phi_c + \sum_k \psi_k \, f^{(k-5)/3} \right) \right], \qquad (4.6)$$

where f is the gravitational wave frequency and t_c and Φ_c are the time and phase at coalescence, respectively. In GR, the PN coefficients ψ_k are given by

$$\psi_k = \frac{3}{128\eta}(\pi M_T)^{(k-5)/3}\alpha_k, \tag{4.7}$$

where $M_T = m_1 + m_2$ is the total mass, $\eta = m_1 m_2 / M_T$ is the symmetric mass ratio and the PN coefficients α_k can be found in equation 3.18 of Buonanno et al. (2009) up to $k = 7$ (3.5PN).

Any deviation in the observed waveform data from the expected PN coefficients could be an indication of a deviation from GR. In order to test for a violation, some freedom is introduced to the parameters so that the data can favour a non-GR waveform if this provides a better fit. Following the method described in Abbott et al. (2016j), the deviation is allowed by introducing additional parameters for each of the coefficients so that

$$\psi_{k,\text{free}} = \psi_{k,\text{gr}} + \Delta\psi_k, \tag{4.8}$$

where $\psi_{k,\text{gr}}$ are the GR values and the sampled parameter is $\delta\chi_k$ so that

$$\psi_{k,\text{free}} = \psi_{k,\text{gr}}(1 + \delta\chi_k) = \psi_{k,\text{gr}}\left(1 + \frac{\Delta\psi_k}{\psi_{k,\text{gr}}}\right). \tag{4.9}$$

In parameter estimation for GW observations, the waveform parameters θ are sampled (either with a nested sampling or MCMC algorithm) in order to produce a posterior distribution for the parameters given the observation data d. With Bayes' theorem, the posterior distribution for θ and the additional parameters $\delta\chi_k$ is given by

$$p(\theta, \delta\chi_k|d) = \frac{p(\theta, \delta\chi_k|I)p(d|h[\theta, \delta\chi_k], I)}{p(d|I)}, \tag{4.10}$$

where $p(\theta, \delta\chi_k|I)$ is the prior for the parameters, $p(d|h[\theta, \delta\chi_k], I)$ the likelihood of the

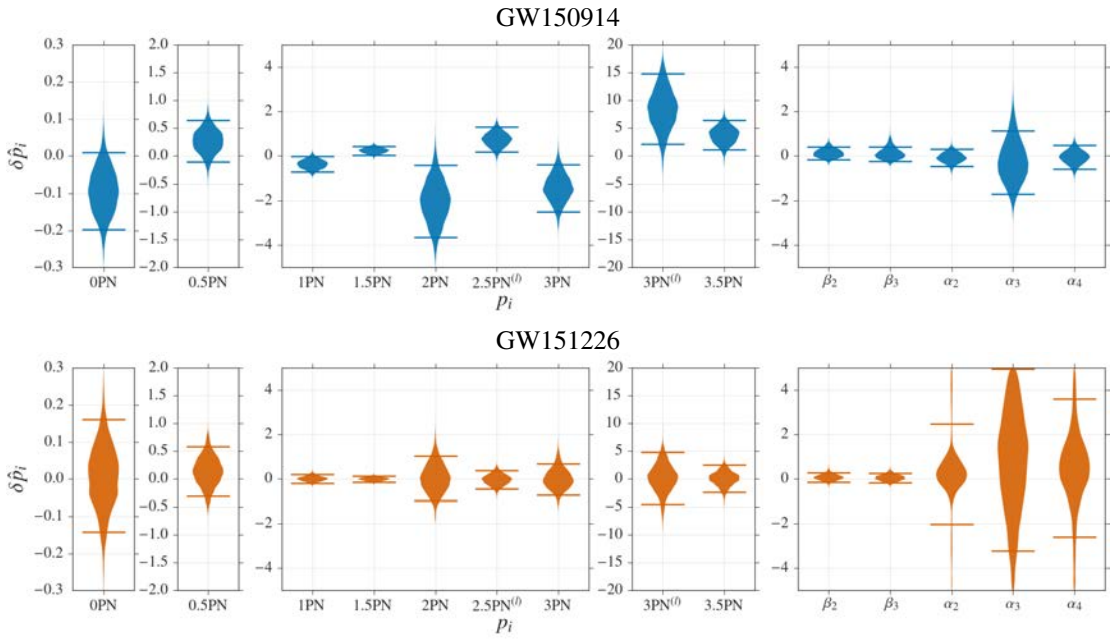


Figure 4.3: Results for tests of general relativity with LIGO observations of binary black hole mergers. The coloured violin bands represent the posterior distribution when including one additional free parameter at a time to search for deviations from general relativity, with zero being the GR value. The lines show the 90% central confidence bands of the distribution. The PN parameters are those described as in the text, whilst the $\beta_{2,3}$ and $\alpha_{2,3,4}$ are phenomenological parameters which modify the intermediate and merger-ringdown of the waveform and are not discussed here. The top panel shows the result from GW150914 and the lower panel from GW151226. All posteriors are consistent with general relativity. For GW150914, it appears that some posteriors lie at the edge of the GR value, however the reason for this is likely down to the much shorter signal length of this observation in comparison to GW151226. Reproduced from Abbott et al. (2016g).

data given the chosen GW waveform (the model) and some choice of the parameters and $p(d|I)$ is the evidence. So far, the results with this method have been consistent with the GR values. Figure 4.3 shows the posterior distribution when allowing for each of the $\delta\chi_k$ inspiral parameters (0PN–3.5PN) to be free one at a time for the observations GW150914 and GW151226. By taking the quantile at which the GR value (zero) is found in each case, we have the summary statistic we need from each experiment.

Results for gravitational wave parametrised tests

We can compute the *p-values* for each observation alone, finding that for GW150914 and GW151226 we calculate *p-values* of 0.99 and 0.06 respectively, giving no indication of inconsistency with the theory. The joint *p-value* when combining the results of both observations gives 0.66, again confirming that so far, there is no inkling of a violation of GR from GW observations.

4.4.2 Newton's constant G

Newton's constant G, giving the magnitude of the force F between two masses $m_{\{1,2\}}$ at separation r ($F = Gm_1m_2/r^2$) has been subject to many tests and measurements. However there are considerable differences in the values that have been measured. Here we briefly summarise the experimental results of G and apply our method.

Measurements of Newton's constant

The Committee on Data for Science and Technology (CODATA) Task Group on Fundamental Constants provide the scientific and technological community with a self-consistent set of international recommended constants (Mohr et al., 2012). The 2010 CODATA (which we label CODATA-10) recommended value for Newton's constant is $G = 6.67384 \pm 0.00080 \times 10^{-11}\mathrm{m}^3\mathrm{s}^{-2}\mathrm{kg}^{-1}$. Many of the experimental results (see

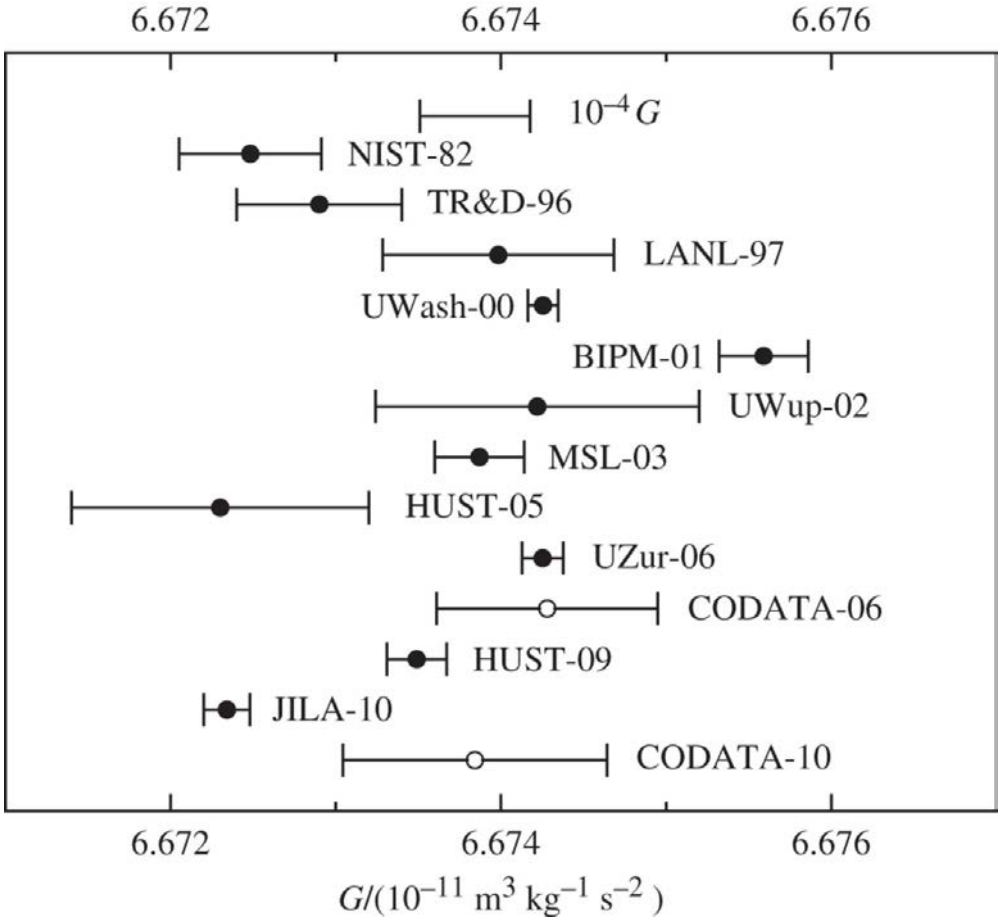


Figure 4.4: The results for Newton's constant included in the CODATA-10 recommendation. The solid circles are input data from experimental results and the clear circles show the CODATA recommended values for 2006 and 2010 (Mohr et al., 2008, 2012). Reproduced from Wood (2014) where there is also listed references to all the experiments included in this figure.

Wood (2014) and references therein) which have been used to produce this value have high precisions of < 50 parts per million, however once collected together the spread of the results is much larger as seen in figure 4.4. The CODATA-10 value itself is produced by multiplying all of the uncertainties for the collected results by some expansion factor in order to increase the uncertainty of the overall recommended value (Wood, 2014).

Applying the method to Newton's constant

In the case of Newton's constant, the null hypothesis is that that the theory is characterised by a given value of G, some G_{theory}. Given a set of experimental measurements of G, which we label $G_{\text{exp},i}$, we make the assumptions that these experiments are independent and not affected by systematic error. By applying our method, we can identify regions of G_{theory} which are inconsistent with the experimental results.

We make the assumption that each experimental result can be approximated as a Gaussian posterior, with values $G_{\text{exp},i}$ and errors $\sigma_{\text{exp},i}$ as shown by figure 4.4,

$$Y_i = \frac{1}{\sigma_{\text{exp},i}\sqrt{2\pi}} \exp\left[-\frac{(G - G_{\text{exp},i})^2}{2\sigma^2_{\text{exp},i}}\right]. \tag{4.11}$$

Our approximations for the posterior distributions assuming this are displayed in the top panel of figure 4.5 (thin-orange lines) along with the CODATA-10 value (thick-black). We then need to choose a theory value, G_{theory}. As we have no theoretical prediction for the value of G_{theory}, here we use the strategy of 'scanning' over possible values of G_{theory} and computing the overall *p-value* for each value scanned over. For each G_{theory}, we calculated the credible interval at which this value is found in the posteriors of the top panel of figure 4.5. As we assume Gaussian posteriors, this is given by

$$p_{\text{cred}} = \text{erf}(r) = \frac{1}{\pi}\int_{-r}^{r} e^{-t^2}\,\text{d}t, \tag{4.12}$$

$$r = \frac{\left|G_{\text{theory}} - G_{\text{exp},i})\right|}{\sqrt{2}\sigma_{\text{exp},i}}, \tag{4.13}$$

giving the credible central region at which G_{theory} is located in the posterior as our *p-value*. We then choose the symmetric *p-value*, p_{symm}

$$p_{\text{symm}} = 2\left(\min\left(p_{\text{cred}},\, 1 - p_{\text{cred}}\right)\right), \tag{4.14}$$

exactly as was done with equation 4.2. For each trial theory value, we now have a list of percentiles from each experiment $p_{\mathrm{symm,i}}(G_{\mathrm{theory}})$ and can conduct our usual test to calculate a degree of inconsistency as described in section 4.2. As the number of experiments here is relatively small $(N = 11)$, this can lessen the power of the test.

Figure 4.5 shows our result for G. The top panel shows our Gaussian estimate of the posterior for each result along with the CODATA-10 value. The centre and lower panels show the result of scanning over theory values of G and calculating the *p-value* at each (the lower panel is identical to the centre but zoomed in the y-axis). We see that either side of the range covered by the experiments, the *p-values* are low as as the consistency is lower at these positions. Within the range covered by the measurements, there is large variation in the calculated *p-values*. The overall consistency is higher around experimental results (where the overall consistency is boosted), and there are notable dips between the experimental values. The lower plot, showing the zoom in between *p-values* of 10^{-5} and 10^{0} shows that there is a considerable dip in consistency on the right-hand side before reaching the highest measured value of G, when the consistency peaks again. These results also show that, as expected by their construction, the CODATA-10 recommendation on the error for G is very conservative, covering the whole of the region with *p-value* $\geq 10^{-5}$.

4.5 Conclusions

Here we have described a generic method of testing the validity of a theory given a set of independent experiments in the case where there is no competing alternative theory. In a contrived situation, we have shown cases where violations can be highlighted with the combination of experiments, whereas individually no one would bat an eyelid at these results. We have applied this method to two situations: the parametrised tests of

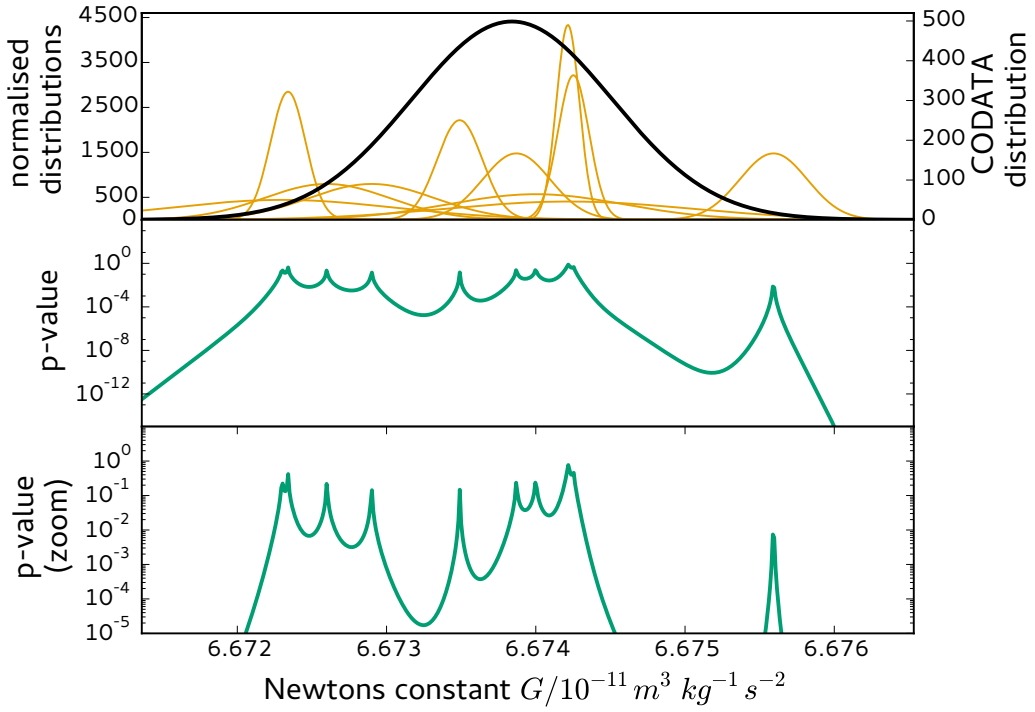


Figure 4.5: Top panel: the orange-thin lines (right axis) show the measurements of G included in the CODATA-10 recommendation assuming Gaussian distributions, whilst the black-thick line (right axis) shows the CODATA-10 value itself. The centre and lower panels show the results of scanning over 'theory' values of G and calculating the *p-value* for consistency at that value. The lower panel, is simply a zoom in of the centre panel in the *p-value* range $\{10^{-5} - 10^{0}\}$.

general relativity with observations of binary black holes, where we find the results to be fully consistent; and also to the experimental measurement of Newton's constant of gravitation, a case in which there is considerable disagreement between measurements.

The method is straightforward, requiring only a *p-value* from each independent experiment, and has the advantage that it does not require joint analysis of data. It also has the attractive quality that tests of different aspects of the same theory can be combined, so that it is very feasible to imagine producing a 'running consistency value' for a theory which is updated as and when new experimental results are released.

Chapter 5

Conclusions

The observation of gravitational waves has opened a new area of astronomy. As more observations are made at different gravitational-wave frequencies, we will continue to build up a picture of the gravitational-wave spectrum, learn about the lives and origins of compact binary systems and find new testing grounds for general relativity. The observation of the gravitational-wave background with pulsar timing arrays will not only confirm that massive black holes do indeed merge, but will also teach us about the massive black hole population and the evolution and lives of galaxies.

Here we summarise our main conclusions from the work presented here and give some brief thoughts to future avenues of investigation.

In Chapter 2 we presented a minimal assumption method to assess what can be learnt about the merging population of massive black hole binaries using pulsar timing array searches for the stochastic gravitational-wave background. Our method makes the simplifying assumption that the binaries are in circular orbits, so that the spectrum can be entirely characterised by a single value. The model we use for the population is extremely generic, being able to reproduce the expected features of the population

whilst not being limited to them. By imposing a prior which allows for a wide range of values, we are able to assess what we can hope to learn about the merging population from pulsar timing array results alone. Our results have shown that in the case of an upper limit, we are able to place an upper limit on the merger rate density. By simulating a future detection we found that the merger rate density can be bounded from both sides, however the mass and redshift distributions are more difficult to constrain in this very agnostic framework. However, the addition of an astrophysical prior may be able to provide more information.

In Chapter 3, we considered a more complex scenario in which we remove the assumption of circular binaries driven by radiation reaction only. We use the model from Chen et al. (2017c), which allows for the resulting binaries to have some eccentricity at the time where radiation reaction becomes the dominant process, and the method from Chen et al. (2017a,b), in which information from the shape of the spectrum is also taken into account. By applying astrophysical priors and using the most recent upper limits from the Parkes Pulsar Timing Array (Shannon et al., 2015), we show that at this stage there is no tension between the upper limits and the astrophysical predictions, indicating that there is, as yet, no need to introduce additional physics into the models. Looking to the future, we show that if no detection is made once pulsar timing arrays reach a factor of ten improved sensitivity, then the most optimistic (highest) predictions will be in trouble.

Looking to the future of this work, we would like to continue to improve this method. For example, we are working on better relating our model to the astrophysical observables so that we will be able to more directly compare information gained from pulsar timing arrays to other observational channels. As more constraining upper limit results (and eventually a detection) are made, we will continue to assess what information can be gleaned from pulsar timing array searches for the gravitational-wave background.

We would also like to investigate the use of this method for other gravitational-wave frequencies, such as the stochastic backgrounds accessible by LISA and LIGO.

Within chapter 3 we considered several astrophysical models which we can compare against each other to make relative statements on the probability of a model given the data. However this scenario, where multiple options are available, is not always the case. In the case where there is no viable alternative to a model, Bayesian model selection cannot be performed. In chapter 4 we describe a method based on the Fisher combined probability test with which we can combine the results of multiple tests of the same theory to make an overall statement in the absence of an alternative theory. Focusing on two cases related to gravity, we apply this method to parametrised tests of gravitational wave observations, finding them to be fully consistent with general relativity and also to the experimental measurement of Newton's constant of gravitation, finding considerable inconsistencies within the range of values that experimental measurements cover. For the future, we intend to combined multiple tests probing different sectors of general relativity, to create an overall validity of the theory. Given the simplicity and generality of this method, we would like to apply it in many situations to produce a running 'validity value' of a theory which can be continually updated as and when new results are made available.

Bibliography

The LIGO Scientific Collaboration J Aasi, B P Abbott, R Abbott, T Abbott, M R Abernathy, K Ackley, C Adams, T Adams, P Addesso, R X Adhikari, and et al. Advanced ligo. *Classical and Quantum Gravity*, 32(7):074001, 2015. URL `http://stacks.iop.org/0264-9381/32/i=7/a=074001`.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observing gravitational-wave transient GW150914 with minimal assumptions. *Phys. Rev. D*, 93(12):122004, June 2016a. doi: 10.1103/PhysRevD.93.122004.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6): 061102, February 2016b. doi: 10.1103/PhysRevLett.116.061102.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Improved Analysis of GW150914 Using a Fully Spin-Precessing Waveform Model. *Physical Review X*, 6 (4):041014, October 2016c. doi: 10.1103/PhysRevX.6.041014.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. GW151226: Observa-

tion of Gravitational Waves from a 22-Solar-Mass Binary Black Hole Coalescence. *Physical Review Letters*, 116(24):241103, June 2016d. doi: 10.1103/PhysRevLett.116.241103.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Comprehensive all-sky search for periodic gravitational waves in the sixth science run LIGO data. *Phys. Rev. D*, 94(4):042002, August 2016e. doi: 10.1103/PhysRevD.94.042002.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. GW150914: Implications for the Stochastic Gravitational-Wave Background from Binary Black Holes. *Physical Review Letters*, 116(13):131102, April 2016f. doi: 10.1103/PhysRevLett.116.131102.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Binary Black Hole Mergers in the First Advanced LIGO Observing Run. *Physical Review X*, 6(4):041015, October 2016g. doi: 10.1103/PhysRevX.6.041015.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Prospects for Observing and Localizing Gravitational-Wave Transients with Advanced LIGO and Advanced Virgo. *Living Reviews in Relativity*, 19:1, February 2016h. doi: 10.1007/lrr-2016-1.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Properties of the Binary Black Hole Merger GW150914. *Physical Review Letters*, 116(24):241102, June 2016i. doi: 10.1103/PhysRevLett.116.241102.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley,

C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Tests of General Relativity with GW150914. *Physical Review Letters*, 116(22):221101, June 2016j. doi: 10.1103/PhysRevLett.116.221101.

B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, K. Ackley, C. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, and et al. Exploring the sensitivity of next generation gravitational wave detectors. *Classical and Quantum Gravity*, 34(4): 044001, February 2017a. doi: 10.1088/1361-6382/aa51f4.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, and et al. GW170104: Observation of a 50-Solar-Mass Binary Black Hole Coalescence at Redshift 0.2. *Physical Review Letters*, 118(22):221101, June 2017b. doi: 10.1103/PhysRevLett.118.221101.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al. First low-frequency Einstein@Home all-sky search for continuous gravitational waves in Advanced LIGO data. *ArXiv e-prints*, July 2017c.

B. P. Abbott, R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, and et al. Upper Limits on Gravitational Waves from Scorpius X-1 from a Model-Based Cross-Correlation Search in Advanced LIGO Data, June 2017d.

R. Abbott, T. D. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. GW170814: A three-detector observation of gravitational waves from a binary black hole coalescence. *ArXiv e-prints*, September 2017e.

B. Abolfathi, D. S. Aguado, G. Aguilar, C. Allende Prieto, A. Almeida, T. Tasnim Ananna, F. Anders, S. F. Anderson, B. H. Andrews, B. Anguiano, and et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data

from the extended Baryon Oscillation Sky Survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. *ArXiv e-prints*, July 2017.

F. Acernese, M. Agathos, K. Agatsuma, D. Aisa, N. Allemandou, A. Allocca, J. Amarni, P. Astone, G. Balestri, G. Ballardin, and et al. Advanced Virgo: a second-generation interferometric gravitational wave detector. *Classical and Quantum Gravity*, 32(2): 024001, January 2015. doi: 10.1088/0264-9381/32/2/024001.

C. Affeldt, K. Danzmann, K. L. Dooley, H. Grote, M. Hewitson, S. Hild, J. Hough, J. Leong, H. Lück, M. Prijatelj, and et al. Advanced techniques in GEO 600. *Classical and Quantum Gravity*, 31(22):224002, November 2014. doi: 10.1088/0264-9381/31/22/224002.

P. Amaro-Seoane, J. R. Gair, M. Freitag, M. C. Miller, I. Mandel, C. J. Cutler, and S. Babak. TOPICAL REVIEW: Intermediate and extreme mass-ratio inspirals — astrophysics, science applications and detection using LISA. *Classical and Quantum Gravity*, 24:R113–R169, September 2007. doi: 10.1088/0264-9381/24/17/R01.

P. Amaro-Seoane, S. Aoudia, H. Audley, G. Auger, S. Babak, J. Baker, E. Barausse, S. Barke, M. Bassan, V. Beckmann, and et al. The Gravitational Universe. *ArXiv e-prints*, May 2013.

P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, and et al. Laser Interferometer Space Antenna. *ArXiv e-prints*, February 2017.

T. An, X. Chen, P. Mohan, and B.-Q. Lao. Radio Frequency Interference Mitigation. *ArXiv e-prints*, November 2017.

N. Andersson, J. Baker, K. Belczynski, S. Bernuzzi, E. Berti, L. Cadonati, P. Cerdá-Durán, J. Clark, M. Favata, L. S. Finn, and et al. The transient gravitational-wave

sky. *Classical and Quantum Gravity*, 30(19):193002, October 2013. doi: 10.1088/0264-9381/30/19/193002.

M. Armano, H. Audley, G. Auger, J. T. Baird, M. Bassan, P. Binetruy, M. Born, D. Bortoluzzi, N. Brandt, M. Caleno, and et al. Sub-femto-g free fall for space-based gravitational wave observatories: Lisa pathfinder results. *Phys. Rev. Lett.*, 116:231101, Jun 2016. doi: 10.1103/PhysRevLett.116.231101. URL `https://link.aps.org/doi/10.1103/PhysRevLett.116.231101`.

Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, J. M. Cordes, P. B. Demorest, X. Deng, T. Dolch, J. A. Ellis, et al., and NANOGrav Collaboration. Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits: Limits from the North American Nanohertz Observatory for Gravitational Waves. *Astrophysical Journal*, 794:141, October 2014. doi: 10.1088/0004-637X/794/2/141.

Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, S. Chatterjee, B. Christy, J. M. Cordes, N. J. Cornish, K. Crowter, P. B. Demorest, et al., and NANOGrav Collaboration. The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background. *Astrophysical Journal*, 821:13, April 2016. doi: 10.3847/0004-637X/821/1/13.

Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto. Interferometer design of the KAGRA gravitational wave detector. *Phys. Rev. D*, 88(4):043007, August 2013. doi: 10.1103/PhysRevD.88.043007.

S. Babak, A. Petiteau, A. Sesana, P. Brem, P. A. Rosado, S. R. Taylor, A. Lassus, J. W. T. Hessels, C. G. Bassa, M. Burgay, and et al. European Pulsar Timing Array Limits on

Continuous Gravitational Waves from Individual Supermassive Black Hole Binaries. *ArXiv e-prints*, September 2015.

S. Babak, J. Gair, A. Sesana, E. Barausse, C. F. Sopuerta, C. P. L. Berry, E. Berti, P. Amaro-Seoane, A. Petiteau, and A. Klein. Science with the space-based interferometer LISA. V. Extreme mass-ratio inspirals. *Phys. Rev. D*, 95(10):103012, May 2017. doi: 10.1103/PhysRevD.95.103012.

D. C. Backer, S. R. Kulkarni, C. Heiles, M. M. Davis, and W. M. Goss. A millisecond pulsar. *Nature*, 300:615–618, December 1982. doi: 10.1038/300615a0.

L. Barack and C. Cutler. LISA capture sources: Approximate waveforms, signal-to-noise ratios, and parameter estimation accuracy. *Phys. Rev. D*, 69(8):082005, April 2004. doi: 10.1103/PhysRevD.69.082005.

M. C. Begelman, R. D. Blandford, and M. J. Rees. Massive black hole binaries in active galactic nuclei. *Nature*, 287:307–309, September 1980. doi: 10.1038/287307a0.

A. Beifiori, S. Courteau, E. M. Corsini, and Y. Zhu. On the correlations between galaxy properties and supermassive black hole mass. *Mon. Not. R. Astron. Soc.*, 419: 2497–2528, January 2012. doi: 10.1111/j.1365-2966.2011.19903.x.

E. F. Bell, S. Phleps, R. S. Somerville, C. Wolf, A. Borch, and K. Meisenheimer. The Merger Rate of Massive Galaxies. *Astrophysical Journal*, 652:270–276, November 2006. doi: 10.1086/508408.

M. Bernardi, A. Meert, R. K. Sheth, M. Huertas-Company, C. Maraston, F. Shankar, and V. Vikram. The massive end of the luminosity and stellar mass functions and clustering from CMASS to SDSS: evidence for and against passive evolution. *MNRAS*, 455:4122–4135, February 2016. doi: 10.1093/mnras/stv2487.

C. P. L. Berry, I. Mandel, H. Middleton, L. P. Singer, A. L. Urban, A. Vecchio, S. Vitale, K. Cannon, B. Farr, W. M. Farr, and et al. Parameter Estimation for Binary Neutron-star Coalescences with Realistic Noise during the Advanced LIGO Era. *Astrophysical Journal*, 804:114, May 2015. doi: 10.1088/0004-637X/804/2/114.

M. Bonetti, F. Haardt, A. Sesana, and E. Barausse. Post-Newtonian evolution of massive black hole triplets in galactic nuclei - I. Numerical implementation and tests. *MNRAS*, 461:4419–4434, October 2016. doi: 10.1093/mnras/stw1590.

M. Bonetti, F. Haardt, A. Sesana, and E. Barausse. Post-Newtonian evolution of massive black hole triplets in galactic nuclei – II. Survey of the parameter space. *ArXiv e-prints*, September 2017a.

M. Bonetti, A. Sesana, E. Barausse, and F. Haardt. Post-Newtonian evolution of massive black hole triplets in galactic nuclei – III. A robust lower limit to the nHz stochastic background of gravitational waves. *ArXiv e-prints*, September 2017b.

R. S. Booth, W. J. G. de Blok, J. L. Jonas, and B. Fanaroff. MeerKAT Key Project Science, Specifications, and Proposals. *ArXiv e-prints*, October 2009.

P. Brook. Emission and rotational variability in pulsars. In *American Astronomical Society Meeting Abstracts*, volume 225 of *American Astronomical Society Meeting Abstracts*, page 421.04, January 2015.

P. Brook, S. Johnston, A. Karastergiou, M. Kerr, and R. Shannon. Correlated torque changes and emission variability in pulsars. ATNF Proposal, April 2014.

K. Bundy, M. Fukugita, R. S. Ellis, T. A. Targett, S. Belli, and T. Kodama. The Greater Impact of Mergers on the Growth of Massive Galaxies: Implications for Mass Assembly and Evolution since z sime 1. *Astrophys. J.*, 697:1369–1383, June 2009. doi: 10.1088/0004-637X/697/2/1369.

A. Buonanno, B. R. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash. Comparison of post-Newtonian templates for compact binary inspiral signals in gravitational-wave detectors. *Phys. Rev. D*, 80(8):084043, October 2009. doi: 10.1103/PhysRevD.80.084043.

M. Burgay, N. D'Amico, A. Possenti, R. N. Manchester, A. G. Lyne, B. C. Joshi, M. A. McLaughlin, M. Kramer, J. M. Sarkissian, F. Camilo, V. Kalogera, C. Kim, and D. R. Lorimer. An increased estimate of the merger rate of double neutron stars from observations of a highly relativistic system. *Nature*, 426:531–533, December 2003. doi: 10.1038/nature02124.

S. Burke-Spolaor. Gravitational-Wave Detection and Astrophysics with Pulsar Timing Arrays. *ArXiv e-prints*, November 2015.

S. Chandrasekhar. Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *Astrophysical Journal*, 97:255, March 1943a. doi: 10.1086/144517.

S. Chandrasekhar. Dynamical Friction. II. The Rate of Escape of Stars from Clusters and the Evidence for the Operation of Dynamical Friction. *Astrophysical Journal*, 97:263, March 1943b. doi: 10.1086/144518.

S. Chen, H. Middleton, A. Sesana, W. Del Pozzo, and A. Vecchio. Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays. *MNRAS*, 468:404–417, June 2017a. doi: 10.1093/mnras/stx475.

S. Chen, H. Middleton, A. Sesana, W. Del Pozzo, and A. Vecchio. Erratum: Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays. *MNRAS*, 469:2455–2456, August 2017b. doi: 10.1093/mnras/stx1064.

S. Chen, A. Sesana, and W. Del Pozzo. Efficient computation of the gravitational wave spectrum emitted by eccentric massive black hole binaries in stellar environments. *MNRAS*, 470:1738–1749, September 2017c. doi: 10.1093/mnras/stx1093.

I. Cognard and D. C. Backer. Timing Noise in PSR 1821-24 : a Micro-Glitch Observed in a Recycled Millisecond Pulsar. In F. A. Rasio and I. H. Stairs, editors, *Binary Radio Pulsars*, volume 328 of *Astronomical Society of the Pacific Conference Series*, page 389, July 2005.

D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365: 11–28, January 2006. doi: 10.1111/j.1365-2966.2005.09675.x.

J. Cuadra, P. J. Armitage, R. D. Alexander, and M. C. Begelman. Massive black hole binary mergers within subparsec scale gas discs. *MNRAS*, 393:1423–1432, March 2009. doi: 10.1111/j.1365-2966.2008.14147.x.

C. Cutler and K. S. Thorne. An Overview of Gravitational-Wave Sources. In N. T. Bishop and S. D. Maharaj, editors, *General Relativity and Gravitation*, pages 72–111, September 2002. doi: 10.1142/9789812776556_0004.

L. de Ravel, O. Le Fèvre, L. Tresse, D. Bottini, B. Garilli, V. Le Brun, D. Maccagni, R. Scaramella, M. Scodeggio, G. Vettolani, and et al. The VIMOS VLT Deep Survey. Evolution of the major merger rate since z ˜ 1 from spectroscopically confirmed galaxy pairs. *A&A*, 498:379–397, May 2009. doi: 10.1051/0004-6361/200810569.

Del Pozzo et al. (in prep.).

P. B. Demorest, R. D. Ferdman, M. E. Gonzalez, D. Nice, S. Ransom, I. H. Stairs, Z. Arzoumanian, A. Brazier, S. Burke-Spolaor, S. J. Chamberlin, and et al. Limits on

the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves. *Astrophysical Journal*, 762:94, January 2013. doi: 10.1088/0004-637X/762/2/94.

S. Detweiler. Pulsar timing measurements and the search for gravitational waves. *Astrophysical Journal*, 234:1100–1104, December 1979. doi: 10.1086/157593.

P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. W. Lazio. The Square Kilometre Array. *IEEE Proceedings*, 97:1482–1496, August 2009. doi: 10.1109/JPROC.2009.2021005.

S. G. Djorgovski, A. J. Drake, A. A. Mahabal, M. J. Graham, C. Donalek, R. Williams, E. C. Beshore, S. M. Larson, J. Prieto, M. Catelan, E. Christensen, and R. H. McNaught. The Catalina Real-Time Transient Survey (CRTS). *ArXiv e-prints*, February 2011.

T. Do, S. A. Wright, A. J. Barth, E. J. Barton, L. Simard, J. E. Larkin, A. M. Moore, L. Wang, and B. Ellerbroek. Prospects for Measuring Supermassive Black Hole Masses with Future Extremely Large Telescopes. *Astronomical Journal*, 147:93, April 2014. doi: 10.1088/0004-6256/147/4/93.

M. Dotti, A. Sesana, and R. Decarli. Massive Black Hole Binaries: Dynamical Evolution and Observational Signatures. *Advances in Astronomy*, 2012:940568, 2012. doi: 10.1155/2012/940568.

M. Dotti, A. Merloni, and C. Montuori. Linking the fate of massive black hole binaries to the active galactic nuclei luminosity function. *MNRAS*, 448:3603–3607, April 2015. doi: 10.1093/mnras/stv291.

A. Einstein. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik*, 354:769–822, 1916. doi: 10.1002/andp.19163540702.

Albert Einstein. ber Gravitationswellen. *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)*, 1918:154–167, 1918.

F. Eisenhauer, R. Genzel, T. Alexander, R. Abuter, T. Paumard, T. Ott, A. Gilbert, S. Gillessen, M. Horrobin, S. Trippe, and et al. SINFONI in the Galactic Center: Young Stars and Infrared Flares in the Central Light-Month. *Astrophysical Journal*, 628:246–259, July 2005. doi: 10.1086/430667.

M. Enoki and M. Nagashima. The Effect of Orbital Eccentricity on Gravitational Wave Background Radiation from Supermassive Black Hole Binaries. *Progress of Theoretical Physics*, 117:241–256, February 2007. doi: 10.1143/PTP.117.241.

S. Fairhurst. Source localization with an advanced gravitational wave detector network. *Classical and Quantum Gravity*, 28(10):105021, May 2011. doi: 10.1088/0264-9381/28/10/105021.

B. Farr, C. P. L. Berry, W. M. Farr, C.-J. Haster, H. Middleton, K. Cannon, P. B. Graff, C. Hanna, I. Mandel, C. Pankow, and et al. Parameter Estimation on Gravitational Waves from Neutron-star Binaries with Spinning Components. *Astrophysical Journal*, 825:116, July 2016. doi: 10.3847/0004-637X/825/2/116.

L. Ferrarese and D. Merritt. A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies. *Astrophysical Journal*, 539:L9–L12, August 2000. doi: 10.1086/312838.

L. S. Finn and D. F. Chernoff. Observing binary inspiral in gravitational radiation: One interferometer. *Phys. Rev. D*, 47:2198–2219, March 1993. doi: 10.1103/PhysRevD.47.2198.

Ronald Fisher. *Statistical methods for research workers*. Oliver and Boyd, Edinburgh, 1970. ISBN 0050021702.

É. É. Flanagan and S. A. Hughes. The basics of gravitational wave theory. *New Journal of Physics*, 7:204, September 2005. doi: 10.1088/1367-2630/7/1/204.

A. R. Foley, T. Alberts, R. P. Armstrong, A. Barta, E. F. Bauermeister, H. Bester, S. Blose, R. S. Booth, D. H. Botha, S. J. Buchner, and et al. Engineering and science highlights of the KAT-7 radio telescope. *MNRAS*, 460:1664–1679, August 2016. doi: 10.1093/mnras/stw1040.

D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC Hammer. *PASP*, 125:306–312, March 2013. doi: 10.1086/670067.

R. S. Foster and D. C. Backer. Constructing a pulsar timing array. *Astrophysical Journal*, 361:300–308, September 1990. doi: 10.1086/169195.

S. M. Gaebel and J. Veitch. How would GW150914 look with future GW detector networks? *ArXiv e-prints*, March 2017.

J. R. Gair, M. Vallisneri, S. L. Larson, and J. G. Baker. Testing General Relativity with Low-Frequency, Space-Based Gravitational-Wave Detectors. *Living Reviews in Relativity*, 16:7, September 2013. doi: 10.12942/lrr-2013-7.

K. Gebhardt, R. Bender, G. Bower, A. Dressler, S. M. Faber, A. V. Filippenko, R. Green, C. Grillmair, L. C. Ho, J. Kormendy, and et al. A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion. *Astrophysical Journal*, 539:L13–L16, August 2000. doi: 10.1086/312840.

A. M. Ghez, S. Salim, N. N. Weinberg, J. R. Lu, T. Do, J. K. Dunn, K. Matthews, M. R. Morris, S. Yelda, E. E. Becklin, and et al. Measuring Distance and Properties of the Milky Way's Central Supermassive Black Hole with Stellar Orbits. *Astrophysical Journal*, 689:1044-1062, December 2008. doi: 10.1086/592738.

S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexander, R. Genzel, F. Martins, and T. Ott. Monitoring Stellar Orbits Around the Massive Black Hole in the Galactic Center. *Astrophysical Journal*, 692:1075–1109, February 2009. doi: 10.1088/0004-637X/692/2/1075.

R. Gilmozzi and J. Spyromilio. The European Extremely Large Telescope (E-ELT). *The Messenger*, 127, March 2007.

A. W. Graham. Breaking the Law: The M $_{bh}$-M $_{spheroid}$ Relations for Core-Sérsic and Sérsic Galaxies. *Astrophys. J.*, 746:113, February 2012. doi: 10.1088/0004-637X/746/1/113.

A. W. Graham and N. Scott. The (black hole mass)-(host spheroid luminosity) relation at high and low masses, the quadratic growth of black holes, and intermediate-mass black hole candidates. *ArXiv e-prints*, November 2012.

A. W. Graham, C. A. Onken, E. Athanassoula, and F. Combes. An expanded M_{bh}-σ diagram, and a new calibration of active galactic nuclei masses. *Mon. Not. R. Astron. Soc.*, 412:2211–2228, April 2011. doi: 10.1111/j.1365-2966.2010.18045.x.

M. J. Graham, S. G. Djorgovski, D. Stern, A. J. Drake, A. A. Mahabal, C. Donalek, E. Glikman, S. Larson, and E. Christensen. A systematic search for close supermassive black hole binaries in the Catalina Real-time Transient Survey. *MNRAS*, 453:1562–1576, October 2015a. doi: 10.1093/mnras/stv1726.

M. J. Graham, S. G. Djorgovski, D. Stern, E. Glikman, A. J. Drake, A. A. Mahabal, C. Donalek, S. Larson, and E. Christensen. A possible close supermassive black-hole binary in a quasar with optical periodicity. *Nature*, 518:74–76, February 2015b. doi: 10.1038/nature14143.

Phil C. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, Department of Physics and Astronomy, University of British Columbia, 2005. ISBN 978-0-521-15012-5. Paperback edition 2010.

K. Gültekin, D. O. Richstone, K. Gebhardt, T. R. Lauer, S. Tremaine, M. C. Aller, R. Bender, A. Dressler, S. M. Faber, A. V. Filippenko, R. Green, L. C. Ho, J. Kormendy, J. Magorrian, J. Pinkney, and C. Siopis. The M-σ and M-L Relations in Galactic Bulges, and Determinations of Their Intrinsic Scatter. *Astrophysical Journal*, 698:198–221, June 2009. doi: 10.1088/0004-637X/698/1/198.

N. Häring and H.-W. Rix. On the Black Hole Mass-Bulge Mass Relation. *Astrophys. J. Letters*, 604:L89–L92, April 2004. doi: 10.1086/383567.

C.-J. Haster, Z. Wang, C. P. L. Berry, S. Stevenson, J. Veitch, and I. Mandel. Inference on gravitational waves from coalescences of stellar-mass compact objects and intermediate-mass black holes. *MNRAS*, 457:4499–4506, April 2016. doi: 10.1093/mnras/stw233.

W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi: 10.1093/biomet/57.1.97. URL `http://biomet.oxfordjournals.org/cgi/content/abstract/57/1/97`.

R. W. Hellings and G. S. Downs. Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *Astrophysical Journal*, 265:L39–L42, February 1983. doi: 10.1086/183954.

L. Hernquist. An analytical model for spherical galaxies and bulges. *Astrophysical Journal*, 356:359–364, June 1990. doi: 10.1086/168845.

A. Hewish, S. J. Bell, J. D. H. Pilkington, P. F. Scott, and R. A. Collins. Observation

of a Rapidly Pulsating Radio Source (Reprinted from Nature, February 24, 1968). *Nature*, 224:472, November 1969. doi: 10.1038/224472b0.

J. Hlavacek-Larrondo, A. C. Fabian, A. C. Edge, and M. T. Hogan. On the hunt for ultramassive black holes in brightest cluster galaxies. *MNRAS*, 424:224–231, July 2012. doi: 10.1111/j.1365-2966.2012.21187.x.

G. Hobbs and S. Dai. A review of pulsar timing array gravitational wave research. *ArXiv e-prints*, July 2017.

G. Hobbs, R. Edwards, and R. Manchester. TEMPO2: a New Pulsar Timing Package. *Chinese Journal of Astronomy and Astrophysics Supplement*, 6(2):189–192, December 2006.

G. Hobbs, A. Archibald, Z. Arzoumanian, D. Backer, M. Bailes, N. D. R. Bhat, M. Burgay, S. Burke-Spolaor, D. Champion, I. Cognard, and et al. The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27(8):084013, April 2010. doi: 10.1088/0264-9381/27/8/084013.

L. Hoffman and A. Loeb. Dynamics of triple black hole systems in hierarchically merging massive galaxies. *MNRAS*, 377:957–976, May 2007. doi: 10.1111/j.1365-2966.2007.11694.x.

D. W. Hogg. Distance measures in cosmology. *ArXiv Astrophysics e-prints*, May 1999.

K. Holley-Bockelmann and F. M. Khan. Galaxy Rotation and Rapid Supermassive Binary Coalescence. *Astrophysical Journal*, 810:139, September 2015. doi: 10.1088/0004-637X/810/2/139.

E. A. Huerta, S. T. McWilliams, J. R. Gair, and S. R. Taylor. Detection of eccentric supermassive black hole binaries with pulsar timing arrays: Signal-to-noise ratio

calculations. *Phys. Rev. D*, 92(6):063010, September 2015. doi: 10.1103/PhysRevD.92.063010.

R. A. Hulse and J. H. Taylor. Discovery of a pulsar in a binary system. *Astrophysical Journal*, 195:L51–L53, January 1975. doi: 10.1086/181708.

O. Ilbert, H. J. McCracken, O. Le Fèvre, P. Capak, J. Dunlop, A. Karim, M. A. Renzini, K. Caputi, S. Boissier, S. Arnouts, and et al. Mass assembly in quiescent and star-forming galaxies since z $\sim$ 4 from UltraVISTA. *A&A*, 556:A55, August 2013. doi: 10.1051/0004-6361/201321100.

B. Iyer, T. Souradeep, C. S. Unnikrishnan, S. Dhurandhar, S. Raja, A. Kumar, and A. (IndIGO Consortium) Sengupta. LIGO-India. Proposal for an interferometric gravitational-wave observatory. *LIGO-DCC M1100296-v2*, 2011. URL `https://dcc.ligo.org/LIGO-M1100296/public`. Accessed August 2017.

A. H. Jaffe and D. C. Backer. Gravitational Waves Probe the Coalescence Rate of Massive Black Hole Binaries. *Astrophysical Journal*, 583:616–631, February 2003. doi: 10.1086/345443.

G. Janssen, G. Hobbs, M. McLaughlin, C. Bassa, A. Deller, M. Kramer, K. Lee, C. Mingarelli, P. Rosado, S. Sanidas, and et al. Gravitational Wave Astronomy with the SKA. *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, art. 37, April 2015.

E T. Jaynes. *Probability Theory, The Logic of Science*. Cambridge University Press, 2003. ISBN 978-0-521-59271-0. Edited by G. Larry Bretthorst, Hardback edition 2009.

F. Jenet, L. S. Finn, J. Lazio, A. Lommen, M. McLaughlin, I. Stairs, D. Stinebring, J. Verbiest, A. Archibald, Z. Arzoumanian, D. Backer, J. Cordes, P. Demorest, R. Fer-

dman, P. Freire, M. Gonzalez, V. Kaspi, V. Kondratiev, D. Lorimer, R. Lynch, D. Nice, S. Ransom, R. Shannon, and X. Siemens. The North American Nanohertz Observatory for Gravitational Waves. *ArXiv e-prints*, September 2009.

F. A. Jenet and J. D. Romano. Understanding the gravitational-wave Hellings and Downs curve for pulsar timing arrays in terms of sound and electromagnetic waves. *American Journal of Physics*, 83:635–645, July 2015. doi: 10.1119/1.4916358.

F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards, A. W. Hotan, J. M. Sarkissian, and S. M. Ord. Upper Bounds on the Low-Frequency Stochastic Gravitational Wave Background from Pulsar Timing Observations: Current Limits and Future Prospects. *The Astrophys. J.*, 653:1571–1576, December 2006. doi: 10.1086/508702.

D. I. Jones. Gravitational waves from rotating strained neutron stars. *Classical and Quantum Gravity*, 19:1255–1265, April 2002. doi: 10.1088/0264-9381/19/7/304.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90:773–795, 1995.

G. Kauffmann and M. Haehnelt. A unified model for the evolution of galaxies and quasars. *MNRAS*, 311:576–588, January 2000. doi: 10.1046/j.1365-8711.2000.03077.x.

M. J. Keith, W. Coles, R. M. Shannon, G. B. Hobbs, R. N. Manchester, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, D. J. Champion, A. Chaudhary, and et al. Measurement and correction of variations in interstellar dispersion in high-precision pulsar timing. *MNRAS*, 429:2161–2174, March 2013. doi: 10.1093/mnras/sts486.

L. Z. Kelley, L. Blecha, and L. Hernquist. Massive black hole binary mergers in

dynamical galactic environments. *MNRAS*, 464:3131–3157, January 2017. doi: 10.1093/mnras/stw2452.

F. M. Khan, A. Just, and D. Merritt. Efficient Merger of Binary Supermassive Black Holes in Merging Galaxies. *Astrophysical Journal*, 732:89, May 2011. doi: 10.1088/0004-637X/732/2/89.

M. G. Kitzbichler and S. D. M. White. A calibration of the relation between the abundance of close galaxy pairs and the rate of galaxy mergers. *Mon. Not. R. Astron. Soc.*, 391:1489–1498, December 2008. doi: 10.1111/j.1365-2966.2008.13873.x.

B. Kocsis and A. Sesana. Gas-driven massive black hole binaries: signatures in the nHz gravitational wave background. *MNRAS*, 411:1467–1479, March 2011. doi: 10.1111/j.1365-2966.2010.17782.x.

J. Kormendy and L. C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *ARA&A*, 51:511–653, August 2013. doi: 10.1146/annurev-astro-082708-101811.

M. Kramer, I. H. Stairs, R. N. Manchester, M. A. McLaughlin, A. G. Lyne, R. D. Ferdman, M. Burgay, D. R. Lorimer, A. Possenti, N. D'Amico, J. M. Sarkissian, G. B. Hobbs, J. E. Reynolds, P. C. C. Freire, and F. Camilo. Tests of General Relativity from Timing the Double Pulsar. *Science*, 314:97–102, October 2006. doi: 10.1126/science.1132305.

A. Kulier, J. P. Ostriker, P. Natarajan, C. N. Lackner, and R. Cen. Understanding Black Hole Mass Assembly via Accretion and Mergers at Late Times in Cosmological Simulations. *Astrophysical Journal*, 799:178, February 2015. doi: 10.1088/0004-637X/799/2/178.

Solomon Kullback. *Information Theory and Statistics*. John Wiley & Sons, 1959.

E. Kun, K. É. Gabányi, M. Karouzos, S. Britzen, and L. Á. Gergely. A spinning supermassive black hole binary model consistent with VLBI observations of the S5 1928+738 jet. *MNRAS*, 445:1370–1382, December 2014. doi: 10.1093/mnras/stu1813.

K. J. Lee, C. G. Bassa, G. H. Janssen, R. Karuppusamy, M. Kramer, K. Liu, D. Perrodin, R. Smits, B. W. Stappers, R. van Haasteren, and L. Lentati. Model-based asymptotically optimal dispersion measure correction for pulsar timing. *MNRAS*, 441:2831–2844, July 2014. doi: 10.1093/mnras/stu664.

L. Lentati, S. R. Taylor, C. M. F. Mingarelli, A. Sesana, S. A. Sanidas, A. Vecchio, R. N. Caballero, K. J. Lee, R. van Haasteren, S. Babak, and et al. European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *MNRAS*, 453:2576–2598, November 2015. doi: 10.1093/mnras/stv1538.

K. Liu, E. F. Keane, K. J. Lee, M. Kramer, J. M. Cordes, and M. B. Purver. Profile-shape stability and phase-jitter analyses of millisecond pulsars. *MNRAS*, 420:361–368, February 2012. doi: 10.1111/j.1365-2966.2011.20041.x.

T. Liu, S. Gezari, S. Heinis, E. A. Magnier, W. S. Burgett, K. Chambers, H. Flewelling, M. Huber, K. W. Hodapp, N. Kaiser, and et al. A Periodically Varying Luminous Quasar at z = 2 from the Pan-STARRS1 Medium Deep Survey: A Candidate Supermassive Black Hole Binary in the Gravitational Wave-driven Regime. *Astrophysical Journal*, 803:L16, April 2015. doi: 10.1088/2041-8205/803/2/L16.

G. Lodato, S. Nayakshin, A. R. King, and J. E. Pringle. Black hole mergers: can gas discs solve the 'final parsec' problem? *MNRAS*, 398:1392–1402, September 2009. doi: 10.1111/j.1365-2966.2009.15179.x.

C. López-Sanjuan, O. Le Fèvre, O. Ilbert, L. A. M. Tasca, C. Bridge, O. Cucciati,

P. Kampczyk, L. Pozzetti, C. K. Xu, C. M. Carollo, and et al. The dominant role of mergers in the size evolution of massive early-type galaxies since z ˜ 1. *A&A*, 548: A7, December 2012. doi: 10.1051/0004-6361/201219085.

D. R. Lorimer. Binary and Millisecond Pulsars. *Living Reviews in Relativity*, 11(1): 8, November 2008. ISSN 1433-8351. doi: 10.12942/lrr-2008-8. URL `https://doi.org/10.12942/lrr-2008-8`.

J. M. Lotz, P. Jonsson, T. J. Cox, and J. R. Primack. The effect of mass ratio on the morphology and time-scales of disc galaxy mergers. *Mon. Not. R. Astron. Soc.*, 404: 575–589, May 2010. doi: 10.1111/j.1365-2966.2010.16268.x.

J. M. Lotz, P. Jonsson, T. J. Cox, D. Croton, J. R. Primack, R. S. Somerville, and K. Stewart. The Major and Minor Galaxy Merger Rates at $z < 1.5$. *Astrophysical Journal*, 742:103, December 2011. doi: 10.1088/0004-637X/742/2/103.

David J. C. Mackay. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, The Pitt Building, Trumpington Street, Cambridge, U.K., 2003. ISBN 0 521 64289 1. Reprinted with corrections 2014.

Michele Maggiore. *Gravitational Waves, Theory and Experiments*, volume 1. Oxford University Press, Département de Physique Théorique, Université de Genève, 2008. ISBN 978-0-19-857075-4. Reprinted 2011.

J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S. M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy, and T. Lauer. The Demography of Massive Dark Objects in Galaxy Centers. *Astronomical Journal*, 115:2285–2305, June 1998. doi: 10.1086/300353.

N. J. McConnell and C.-P. Ma. Revisiting the Scaling Relations of Black Hole Masses

and Host Galaxy Properties. *Astrophysical Journal*, 764:184, February 2013. doi: 10.1088/0004-637X/764/2/184.

N. J. McConnell, C.-P. Ma, K. Gebhardt, S. A. Wright, J. D. Murphy, T. R. Lauer, J. R. Graham, and D. O. Richstone. Two ten-billion-solar-mass black holes at the centres of giant elliptical galaxies. *Nature*, 480:215–218, December 2011. doi: 10.1038/nature10636.

J. W. McKee, G. H. Janssen, B. W. Stappers, A. G. Lyne, R. N. Caballero, L. Lentati, G. Desvignes, A. Jessner, C. A. Jordan, R. Karuppusamy, M. Kramer, I. Cognard, D. J. Champion, E. Graikou, P. Lazarus, S. Osłowski, D. Perrodin, G. Shaifullah, C. Tiburzi, and J. P. W. Verbiest. A glitch in the millisecond pulsar J0613-0200. *MNRAS*, 461:2809–2817, September 2016. doi: 10.1093/mnras/stw1442.

S. T. McWilliams, J. P. Ostriker, and F. Pretorius. Gravitational Waves and Stalled Satellites from Massive Galaxy Mergers at z <= 1. *Astrophysical Journal*, 789:156, July 2014. doi: 10.1088/0004-637X/789/2/156.

D. Merritt. Loss-cone dynamics. *Classical and Quantum Gravity*, 30(24):244005, December 2013. doi: 10.1088/0264-9381/30/24/244005.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21:1087–1092, June 1953. doi: 10.1063/1.1699114.

M. Mezcua. Observational evidence for intermediate-mass black holes. *ArXiv e-prints*, May 2017.

H. Middleton, W. Del Pozzo, W. M. Farr, A. Sesana, and A. Vecchio. Astrophysical constraints on massive black hole binary evolution from pulsar timing arrays. *MNRAS*, 455:L72–L76, January 2016. doi: 10.1093/mnrasl/slv150.

H. Middleton, S. Chen, W. Del Pozzo, A. Sesana, and A. Vecchio. No tension between assembly models of supermassive black hole binaries and pulsar observations. *ArXiv e-prints*, July 2017.

S. Mikkola and M. J. Valtonen. Evolution of binaries in the field of light particles and the problem of two black holes. *MNRAS*, 259:115–120, November 1992. doi: 10.1093/mnras/259.1.115.

M. Milosavljević and D. Merritt. The Final Parsec Problem. In J. M. Centrella, editor, *The Astrophysics of Gravitational Wave Sources*, volume 686 of *American Institute of Physics Conference Series*, pages 201–210, October 2003. doi: 10.1063/1.1629432.

C. M. F. Mingarelli, T. Sidery, I. Mandel, and A. Vecchio. Characterizing gravitational wave stochastic background anisotropy with pulsar timing arrays. *Phys. Rev. D*, 88 (6):062005, September 2013. doi: 10.1103/PhysRevD.88.062005.

P. J. Mohr, B. N. Taylor, and D. B. Newell. CODATA recommended values of the fundamental physical constants: 2006. *Reviews of Modern Physics*, 80:633–730, April 2008. doi: 10.1103/RevModPhys.80.633.

P. J. Mohr, B. N. Taylor, and D. B. Newell. CODATA recommended values of the fundamental physical constants: 2010. *Reviews of Modern Physics*, 84:1527–1605, October 2012. doi: 10.1103/RevModPhys.84.1527.

C. J. Moore, R. H. Cole, and C. P. L. Berry. Gravitational-wave sensitivity curves. *Classical and Quantum Gravity*, 32(1):015014, January 2015a. doi: 10.1088/0264-9381/32/1/015014.

C. J. Moore, S. R. Taylor, and J. R. Gair. Estimating the sensitivity of pulsar timing arrays. *Classical and Quantum Gravity*, 32(5):055004, March 2015b. doi: 10.1088/0264-9381/32/5/055004.

C. J. Mundy, C. J. Conselice, K. J. Duncan, O. Almaini, B. Häußler, and W. G. Hartley. A consistent measure of the merger histories of massive galaxies using close-pair statistics - I. Major mergers at $z < 3.5$. *MNRAS*, 470:3507–3531, September 2017. doi: 10.1093/mnras/stx1238.

A. Muzzin, D. Marchesini, M. Stefanon, M. Franx, H. J. McCracken, B. Milvang-Jensen, J. S. Dunlop, J. P. U. Fynbo, G. Brammer, I. Labbé, and P. G. van Dokkum. The Evolution of the Stellar Mass Functions of Star-forming and Quiescent Galaxies to z = 4 from the COSMOS/UltraVISTA Survey. *Astrophysical Journal*, 777:18, November 2013. doi: 10.1088/0004-637X/777/1/18.

R. Nan, D. Li, C. Jin, Q. Wang, L. Zhu, W. Zhu, H. Zhang, Y. Yue, and L. Qian. The Five-Hundred Aperture Spherical Radio Telescope (fast) Project. *International Journal of Modern Physics D*, 20:989–1024, 2011. doi: 10.1142/S0218271811019335.

C. J. Nixon, A. R. King, and J. E. Pringle. The final parsec problem: aligning a binary with an external accretion disc. *MNRAS*, 417:L66–L69, October 2011. doi: 10.1111/j.1745-3933.2011.01121.x.

I. Pâris, P. Petitjean, N. P. Ross, A. D. Myers, É. Aubourg, A. Streblyanska, S. Bailey, É. Armengaud, N. Palanque-Delabrouille, C. Yèche, and et al. The Sloan Digital Sky Survey Quasar Catalog: Twelfth data release. *A&A*, 597:A79, January 2017. doi: 10.1051/0004-6361/201527999.

D. Perrodin, F. Jenet, A. Lommen, L. Finn, P. Demorest, R. Ferdman, M. Gonzalez, D. Nice, S. Ransom, and I. Stairs. Timing Noise Analysis of NANOGrav Pulsars. *ArXiv e-prints*, November 2013.

P. C. Peters and J. Mathews. Gravitational Radiation from Point Masses in a Keplerian Orbit. *Physical Review*, 131:435–440, July 1963. doi: 10.1103/PhysRev.131.435.

E. S. Phinney. A Practical Theorem on Gravitational Wave Backgrounds. *ArXiv Astrophysics e-prints*, August 2001.

PINT. Pint is not tempo3 – new software for high-precision pulsar timing, 2017. URL `https://github.com/nanograv/PINT`.

M. Preto, I. Berentzen, P. Berczik, and R. Spurzem. Fast Coalescence of Massive Black Hole Binaries from Mergers of Galactic Nuclei: Implications for Low-frequency Gravitational-wave Astrophysics. *Astrophysical Journal*, 732:L26, May 2011. doi: 10.1088/2041-8205/732/2/L26.

G. D. Quinlan. The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background. *New Astronomy*, 1:35–56, July 1996. doi: 10.1016/S1384-1076(96)00003-6.

M. Rajagopal and R. W. Romani. Ultra–Low-Frequency Gravitational Radiation from Massive Black Hole Binaries. *Astrophysical Journal*, 446:543, June 1995. doi: 10.1086/175813.

A. Rasskazov and D. Merritt. Evolution Of Massive Black Hole Binaries In Rotating Stellar Nuclei: Implications For Gravitational Wave Detection. *ArXiv e-prints*, June 2016.

V. Ravi, J. S. B. Wyithe, G. Hobbs, R. M. Shannon, R. N. Manchester, D. R. B. Yardley, and M. J. Keith. Does a "Stochastic" Background of Gravitational Waves Exist in the Pulsar Timing Band? *Astrophysical Journal*, 761:84, December 2012. doi: 10.1088/0004-637X/761/2/84.

V. Ravi, J. S. B. Wyithe, R. M. Shannon, and G. Hobbs. Prospects for gravitational-wave detection and supermassive black hole astrophysics with pulsar timing arrays. *MNRAS*, 447:2772–2783, March 2015. doi: 10.1093/mnras/stu2659.

C. Rodriguez, G. B. Taylor, R. T. Zavala, A. B. Peck, L. K. Pollack, and R. W. Romani. A Compact Supermassive Binary Black Hole System. *Astrophysical Journal*, 646: 49–60, July 2006. doi: 10.1086/504825.

L. Sampson, N. J. Cornish, and S. T. McWilliams. Constraining the solution to the last parsec problem with pulsar timing. *Phys. Rev. D*, 91(8):084055, April 2015. doi: 10.1103/PhysRevD.91.084055.

G. H. Sanders. The Thirty Meter Telescope (TMT): An International Observatory. *Journal of Astrophysics and Astronomy*, 34:81–86, June 2013. doi: 10.1007/s12036-013-9169-5.

E. Sani, A. Marconi, L. K. Hunt, and G. Risaliti. The Spitzer/IRAC view of black hole-bulge scaling relations. *Mon. Not. R. Astron. Soc.*, 413:1479–1494, May 2011. doi: 10.1111/j.1365-2966.2011.18229.x.

B. Sathyaprakash, M. Abernathy, F. Acernese, P. Ajith, B. Allen, P. Amaro-Seoane, N. Andersson, S. Aoudia, K. Arun, P. Astone, and et al. Scientific objectives of Einstein Telescope. *Classical and Quantum Gravity*, 29(12):124013, June 2012. doi: 10.1088/0264-9381/29/12/124013.

M. V. Sazhin. Opportunities for detecting ultralong gravitational waves. *Soviet Ast.*, 22: 36–38, February 1978.

J. D. Schnittman and J. H. Krolik. Evolution of a Binary Black Hole with a Retrograde Circumbinary Accretion Disk. *Astrophysical Journal*, 806:88, June 2015. doi: 10.1088/0004-637X/806/1/88.

A. Sesana. Insights into the astrophysics of supermassive black hole binaries from pulsar timing observations. *Classical and Quantum Gravity*, 30(22):224014, November 2013a. doi: 10.1088/0264-9381/30/22/224014.

A. Sesana. Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band. *MNRAS*, 433:L1–L5, June 2013b. doi: 10.1093/mnrasl/slt034.

A. Sesana and F. M. Khan. Scattering experiments meet N-body - I. A practical recipe for the evolution of massive black hole binaries in stellar environments. *MNRAS*, 454: L66–L70, November 2015. doi: 10.1093/mnrasl/slv131.

A. Sesana, A. Vecchio, and C. N. Colacino. The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. *MNRAS*, 390:192–209, October 2008. doi: 10.1111/j.1365-2966.2008.13682.x.

A. Sesana, A. Vecchio, and M. Volonteri. Gravitational waves from resolvable massive black hole binary systems and observations with Pulsar Timing Arrays. *MNRAS*, 394: 2255–2265, April 2009. doi: 10.1111/j.1365-2966.2009.14499.x.

A. Sesana, F. Shankar, M. Bernardi, and R. K. Sheth. Selection bias in dynamically measured supermassive black hole samples: consequences for pulsar timing arrays. *MNRAS*, 463:L6–L11, November 2016. doi: 10.1093/mnrasl/slw139.

F. Shankar, M. Bernardi, R. K. Sheth, L. Ferrarese, A. W. Graham, G. Savorgnan, V. Allevato, A. Marconi, R. Läsker, and A. Lapi. Selection bias in dynamically measured supermassive black hole samples: its consequences and the quest for the most fundamental relation. *MNRAS*, 460:3119–3142, August 2016. doi: 10.1093/mnras/stw678.

R. M. Shannon, V. Ravi, W. A. Coles, G. Hobbs, M. J. Keith, R. N. Manchester, J. S. B. Wyithe, M. Bailes, N. D. R. Bhat, S. Burke-Spolaor, and et al. Gravitational-wave

limits from pulsar timing constrain supermassive black hole evolution. *Science*, 342: 334–337, October 2013.

R. M. Shannon, S. Osłowski, S. Dai, M. Bailes, G. Hobbs, R. N. Manchester, W. van Straten, C. A. Raithel, V. Ravi, L. Toomey, and et al. Limitations in timing precision due to single-pulse shape variability in millisecond pulsars. *MNRAS*, 443:1463–1481, September 2014. doi: 10.1093/mnras/stu1213.

R. M. Shannon, V. Ravi, L. T. Lentati, P. D. Lasky, G. Hobbs, M. Kerr, R. N. Manchester, W. A. Coles, Y. Levin, M. Bailes, and et al. Gravitational waves from binary supermassive black holes missing in pulsar observations. *Science*, 349:1522–1525, September 2015. doi: 10.1126/science.aab1910.

Stuart L. Shapiro and Saul A. Teukolsky. *Black Holes, White Dwarfs and Neutron Stars*. John Wiley & Sons, Cornell University, Ithaca, New York, 1983. ISBN 0-471-87317-9.

X. Siemens, J. Ellis, F. Jenet, and J. D. Romano. The stochastic background: scaling laws and time to detection for pulsar timing arrays. *Classical and Quantum Gravity*, 30(22):224015, November 2013. doi: 10.1088/0264-9381/30/22/224015.

J. Simon and S. Burke-Spolaor. Constraints on Black Hole/Host Galaxy Co-evolution and Binary Stalling Using Pulsar Timing Arrays. *Astrophysical Journal*, 826:11, July 2016. doi: 10.3847/0004-637X/826/1/11.

J Skilling. American institute of physics conference series. In *American Institute of Physics Conference Series*, volume 735, pages 395–405, 2004.

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, and et al. Simulations of the formation, evolution

and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. doi: 10.1038/nature03597.

J. H. Taylor and J. M. Weisberg. Further experimental tests of relativistic gravity using the binary pulsar PSR 1913 + 16. *Astrophysical Journal*, 345:434–450, October 1989. doi: 10.1086/167917.

S. R. Taylor, C. M. F. Mingarelli, J. R. Gair, A. Sesana, G. Theureau, S. Babak, C. G. Bassa, P. Brem, M. Burgay, R. N. Caballero, et al., and EPTA Collaboration. Limits on Anisotropy in the Nanohertz Stochastic Gravitational Wave Background. *Physical Review Letters*, 115(4):041101, July 2015. doi: 10.1103/PhysRevLett.115.041101.

S. R. Taylor, M. Vallisneri, J. A. Ellis, C. M. F. Mingarelli, T. J. W. Lazio, and R. van Haasteren. Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits. *Astrophysical Journal*, 819:L6, March 2016. doi: 10.3847/2041-8205/819/1/L6.

S. R. Taylor, L. Lentati, S. Babak, P. Brem, J. R. Gair, A. Sesana, and A. Vecchio. All correlations must die: Assessing the significance of a stochastic gravitational-wave background in pulsar timing arrays. *Phys. Rev. D*, 95(4):042002, February 2017. doi: 10.1103/PhysRevD.95.042002.

A. R. Tomczak, R. F. Quadri, K.-V. H. Tran, I. Labbé, C. M. S. Straatman, C. Papovich, K. Glazebrook, R. Allen, G. B. Brammer, G. G. Kacprzak, and et al. Galaxy Stellar Mass Functions from ZFOURGE/CANDELS: An Excess of Low-mass Galaxies since z = 2 and the Rapid Buildup of Quiescent Galaxies. *Astrophysical Journal*, 783:85, March 2014. doi: 10.1088/0004-637X/783/2/85.

A. D. Tubbs. The dynamical evolution of NGC 5128. *Astrophysical Journal*, 241: 969–980, November 1980. doi: 10.1086/158411.

M. J. Valtonen, H. J. Lehto, K. Nilsson, J. Heidt, L. O. Takalo, A. Sillanpää, C. Villforth, M. Kidger, G. Poyner, T. Pursimo, and et al. A massive binary black-hole system in OJ287 and a test of general relativity. *Nature*, 452:851–853, April 2008. doi: 10.1038/nature06896.

M. J. Valtonen, S. Ciprini, and H. J. Lehto. On the masses of OJ287 black holes. *MNRAS*, 427:77–83, November 2012. doi: 10.1111/j.1365-2966.2012.21861.x.

R. van Haasteren, Y. Levin, G. H. Janssen, K. Lazaridis, M. Kramer, B. W. Stappers, G. Desvignes, M. B. Purver, A. G. Lyne, R. D. Ferdman, and et al. Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data. *MNRAS*, 414:3117–3128, July 2011. doi: 10.1111/j.1365-2966.2011.18613.x.

R. van Haasteren et al. Erratum: Placing limits on the stochastic gravitational-wave background using European Pulsar Timing Array data. *MNRAS*, 425:1597–1597, September 2012. doi: 10.1111/j.1365-2966.2012.20916.x.

E. Vasiliev, F. Antonini, and D. Merritt. The Final-parsec Problem in Nonspherical Galaxies Revisited. *Astrophysical Journal*, 785:163, April 2014. doi: 10.1088/0004-637X/785/2/163.

E. Vasiliev, F. Antonini, and D. Merritt. The Final-parsec Problem in the Collisionless Limit. *Astrophysical Journal*, 810:49, September 2015. doi: 10.1088/0004-637X/810/1/49.

S. Vaughan, P. Uttley, A. G. Markowitz, D. Huppenkothen, M. J. Middleton, W. N. Alston, J. D. Scargle, and W. M. Farr. False periodicities in quasar time-domain surveys. *MNRAS*, 461:3145–3152, September 2016. doi: 10.1093/mnras/stw1412.

J. Veitch and A. Vecchio. Bayesian coherent analysis of in-spiral gravitational wave

signals with a detector network. *Phys. Rev. D*, 81(6):062003, March 2010. doi: 10.1103/PhysRevD.81.062003.

John Veitch, Walter Del Pozzo, Cody, Matt Pitkin, and ed1d1a8d. johnveitch/cpnest: Minor optimisation, July 2017. URL `https://doi.org/10.5281/zenodo.835874`.

J. P. W. Verbiest, L. Lentati, G. Hobbs, R. van Haasteren, P. B. Demorest, G. H. Janssen, J.-B. Wang, G. Desvignes, R. N. Caballero, M. J. Keith, and et al. The International Pulsar Timing Array: First data release. *MNRAS*, 458:1267–1288, May 2016. doi: 10.1093/mnras/stw347.

M. Volonteri. The Formation and Evolution of Massive Black Holes. *Science*, 337:544–, August 2012. doi: 10.1126/science.1220843.

M. Volonteri, F. Haardt, and P. Madau. The Assembly and Merging History of Supermassive Black Holes in Hierarchical Models of Galaxy Formation. *Astrophysical Journal*, 582:559–573, January 2003. doi: 10.1086/344675.

J. M. Weisberg and J. H. Taylor. The Relativistic Binary Pulsar B1913+16: Thirty Years of Observations and Analysis. In F. A. Rasio and I. H. Stairs, editors, *Binary Radio Pulsars*, volume 328 of *Astronomical Society of the Pacific Conference Series*, page 25, July 2005.

S. D. M. White and M. J. Rees. Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. *MNRAS*, 183:341–358, May 1978.

Barry M. Wood. Recommending a value for the newtonian gravitational constant. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 372(2026), 2014. ISSN 1364-503X. doi: 10.1098/rsta.2014.0029. URL `http://rsta.royalsocietypublishing.org/content/372/2026/20140029`.

J. S. B. Wyithe and A. Loeb. Low-Frequency Gravitational Waves from Massive Black Hole Binaries: Predictions for LISA and Pulsar Timing Arrays. *Astrophysical Journal*, 590:691–706, June 2003. doi: 10.1086/375187.

C. K. Xu, Y. Zhao, N. Scoville, P. Capak, N. Drory, and Y. Gao. Major-merger Galaxy Pairs in the COSMOS Field — Mass-dependent Merger Rate Evolution since z = 1. *Astrophys. J.*, 747:85, March 2012. doi: 10.1088/0004-637X/747/2/85.

Q. Xu, H. Wen, N. Wang, and L. Li. Development challenges for the Xinjiang 110 m radio telescope (QTT) high accuracy panel structures. *Scientia Sinica Physica, Mechanica & Astronomica*, 47(5):059502, May 2017. doi: 10.1360/SSPMA2016-00501.

Q. Yu. Evolution of massive binary black holes. *MNRAS*, 331:935–958, April 2002. doi: 10.1046/j.1365-8711.2002.05242.x.

Appendix A

Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays & Erratum

Here we reproduce, in its published form, the work from Chen et al. (2017a) and Chen et al. (2017b) (erratum).

Authorship:

Siyuan Chen, Hannah Middleton, Walter Del Pozzo, Alberto Sesana, Alberto Vecchio

MNRAS 000, 1–14 (2016) Preprint 8 December 2016 Compiled using MNRAS LATEX style file v3.0

Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays

Siyuan Chen,[1]⋆ Hannah Middleton,[1]† Alberto Sesana,[1] Walter Del Pozzo[1,2] and Alberto Vecchio[1]
[1]School of Physics & Astronomy, University of Birmingham, Birmingham, B15 2TT, UK
[2]Dipartimento di Fisica "Enrico Fermi", Università di Pisa, Pisa I-56127, Italy

Accepted . . . Received . . . ; in original form . . .

ABSTRACT
We consider the inverse problem in pulsar timing array (PTA) analysis, investigating what astrophysical information about the underlying massive black hole binary (MBHB) population can be recovered from the detection of a stochastic gravitational wave background (GWB). We employ a physically motivated model that connects the GWB spectrum to a series of parameters describing the underlying redshift evolution of the MBHB mass function and to the typical eccentricity they acquire while interacting with the dense environment of post merger galactic nuclei. This allows the folding in of information about the spectral shape of the GWB into the analysis. The priors on the model parameters are assumed to be uninformative and consistent with the current lack of secure observations of sub-parsec MBHBs. We explore the implications of current upper limits as well as of future detections with a variety of PTA configurations. We confirm our previous finding that current upper limits can only place an upper bound on the overall MBHB merger rate. Depending on the properties of the array, future detections can also constrain several MBHB population models at different degrees of fidelity. In particular, a simultaneous detection of a steepening of the spectrum at high frequency and a bending at low frequency will place strong constraints on both the MBHB mass function and on the typical eccentricity of inspiralling MBHBs, providing insights on MBHB astrophysics unlikely to be achievable by any other means.

Key words: gravitational waves – black hole physics – pulsars: general – methods: data analysis

1 INTRODUCTION

Massive black holes (MBHs) appear to be a fundamental component in galaxy formation and evolution. In fact, all massive galaxies appear to host MBHs in their centres (Kormendy & Ho 2013, and references therein). In the hierarchical clustering model of structure formation (White & Rees 1978), these MBHs are the dormant counterparts of quasars and active galactic nuclei (e.g. Hopkins et al. 2006). In a nutshell, galaxies grow through a sequence of mergers and accretion episodes that trigger star formation and fuel the central MBHs. Gas accretion powers luminous electromagnetic radiation, which is at the basis of the Quasar phenomenon (e.g. Croton et al. 2006). If most galaxies host MBHs then, following galaxy mergers, the two MBHs sink to the center of the merger remnant eventually forming a bound MBH binary (MBHB Begelman et al. 1980). The details of this general picture are not well understood. In particular, it is not clear whether MBHBs efficiently merge as a consequence of galaxy mergers and what the details of the dynamical processes driving their final coalescence are (see Dotti et al. 2012, and references therein).

MBHBs are among the loudest sources of gravitational waves (GWs) in the Universe, and during their inspiral emit radiation that falls in the nHz frequency range, probed by ongoing and upcoming pulsar timing array (PTA) experiments (Sesana et al. 2008). In fact, GWs imprint a distinctive signature in the time of arrivals (ToAs) of ultra-stable millisecond pulsars (MSPs). This signature can be disentangled from other noise sources by cross-correlating ToA time-series from an ensemble of pulsars (Hellings & Downs 1983). PTAs therefore monitor a large number of MSPs, looking for this distinctive correlation (Foster & Backer 1990). This challenge is currently undertaken by the European Pulsar Timing Array (EPTA Desvignes et al. 2016), the Parkes Pulsar Timing Array (PPTA Reardon et al. 2016) and North American Nanohertz Observatory for Gravitational Waves (NANOGrav The NANOGrav Collaboration et al. 2015). The three collaborations are joining forces under the aegis of the International Pulsar Timing Array (IPTA Verbiest et al. 2016), paving the way towards a future global collaboration that

⋆ E-mail: schen@star.sr.bham.ac.uk
† E-mail: hannahm@star.sr.bham.ac.uk

will take advantage of upcoming facilities such as the South African telescope array MeerKAT (Booth et al. 2009), the Chinese 500-mt telescope FAST (Nan et al. 2011) and eventually the Square Kilometre Array (SKA).

Since PTAs observe individual pulsars with cadence Δt of order of few weeks for an experiment duration T of several years, they are most sensitive to GWs in the frequency range $1/\Delta t < f < T$, i.e. $10^{-9} - 10^{-7}$ Hz. At such low frequencies, the superposition of GW signals emitted by a cosmological population of MBHBs results in a stochastic GW background (GWB), although, especially at high frequencies, particularly massive/nearby systems may be resolved individually (Sesana et al. 2009). The GWB amplitude and spectral shape depends on the underlying population of MBHBs and can therefore be used to constrain their astrophysical and cosmological properties (Sesana 2013a).

Direct detection of GWs by Advanced LIGO recently opened the high frequency GW window on the Universe (Abbott et al. 2016a,c). Besides the profound implications for gravity theory and fundamental physics (Abbott et al. 2016b), from an astrophysical perspective, GWs are a new tool to understand the physics of compact objects populating the Universe, and how they connect with the evolution of gas, stars and galaxies. From this point of view, PTAs provide a formidable tool to understand the dynamics of MBHBs and their demographic along the cosmic history. In fact, the amplitude of the GWB depends on how frequently MBHBs merge and what their typical mass is, whereas the spectral shape also depends on the mechanism driving the MBHB inspiral and, crucially, on their eccentricity. It is well known that under the assumption of circular GW driven binaries, the characteristic GWB strains follows a power-law $h_c \propto f^{-2/3}$ (Phinney 2001). However, at large orbital separations (i.e. at low frequencies), MBHB evolution is dominated by energy and angular momentum exchange with the stellar and gas rich surroundings, potentially growing the MBHB eccentricity and resulting in a low frequency turnover of the GWB (Enoki & Nagashima 2007; Kocsis & Sesana 2011; Sesana 2013a; Ravi et al. 2014; Kelley et al. 2016; Rasskazov & Merritt 2016). Therefore, the characterization of the amplitude and spectral shape of the GWB carries precious information on the underlying population of MBHBs. To what extent such information can be recovered via PTA observations is the main focus of this paper.

PTA's effort has been so far focusing on delivering the best possible ToA datasets (e.g. Verbiest et al. 2016) and on developing the necessary data analysis tools for detection of either a GWB or individual sources (e.g. Ellis et al. 2012; Petiteau et al. 2013; Lentati et al. 2013). The application of the latter to the former resulted so far in upper limits only (Lentati et al. 2015; Babak et al. 2016; Arzoumanian et al. 2016), and in the absence of a detection, little effort has been spent in the 'inverse problem', namely on investigating what astrophysical information can be recovered from PTA observations. This does not mean that astrophysics has been so far ignored; for example, Arzoumanian et al. (2016) discussed in length the consequences of their upper limit for MBHB dynamics, and Simon & Burke-Spolaor (2016) explored the implications for MBH mass-galaxy relations proposed in the literature. However, although astrophysical inference has been applied to specific upper limits, a framework that connects PTA observations to MBHB astrophysics in the general context of any PTA detection is missing. As part of the common effort of the EPTA collaboration (Desvignes et al. 2016) to detect GWs with pulsar timing, this paper is an attempt of making a step forward towards the creation of such a framework.[1]

We consider the model developed in Chen et al. (2016), hereinafter PaperI, for the GWB emitted by a generic population of eccentric MBHBs evolving via scattering of ambient stars. In our model, MBHBs hold a constant eccentricity so long as their evolution is driven by stellar scattering, and circularize under the effect of GW radiation when their dynamics is GW driven (i.e., after decoupling from the stellar environment). In PaperI we showed that the decoupling radius is only a mild fraction of the density of ambient stars, and for stellar density typical of massive galaxies, occurs at frequencies well below the relevant PTA range. As such, we found that the effect of eccentricity is much more prominent, therefore the GWB shape can be fully characterized by a few parameters defining the mass function of MBHBs and its redshift evolution, and the typical eccentricity at decoupling. Expanding on Middleton et al. (2016) (M16 hereinafter), we simulate GWB detection for a variety of PTAs and we investigate to what extent the underlying MBHB population parameters can be constrained.

The paper is organized as follows. In section 2 we summarize the relevant features of the GWB spectral models developed in PaperI. In section 3 we introduce the theory of GWB detection with PTAs and define the impact of the relevant array quantities on the signal-to-noise ratio (S/N) of the measurement. The set-up of our simulations is outlined in section 4 and the analysis method used for astrophysical inference is described in section 5. We present and discuss in detail our results in section 6 and conclude with some final remarks and prospects for future expansion of this work in section 7.

2 ASTROPHYSICAL MODEL

We use the model developed in PaperI for a population of eccentric MBHBs evolving via three-body scattering against the stellar environment. In PaperI, we expressed the properties of the environment (stellar density, velocity dispersion etc.) as a function of the MBHB total mass only; therefore, the MBHB mass defines the relevant stellar background properties, which we take to be consistent with that typical of elliptical galaxies (where the most massive binaries, dominating the GWB, reside). In a nutshell, the stellar density is modelled with a Dehnen profile (Dehnen 1993) with total mass set by the intrinsic relation between the MBH and the galaxy bulge masses – usually referred to as $M_{\rm BH} - M_{\rm bulge}$– provided in Kormendy & Ho (2013), scale radius a defined by the empirical $M_{\rm bulge} - a$ relation found by Dabringhausen et al. (2008)[2], and inner profile slope $\gamma = 1$, appropriate for massive ellipticals. In this model, binaries decouple from the stellar environment at orbital frequencies much lower than the relevant PTA window (which is $f > 1$nHz) and the PTA signal can be constructed taking into account the post-decoupling GW-driven evolution of the eccentric binary only (see PaperI for a full description of the model). The overall GWB spectrum can

[1] While completing this work we became aware of an independent parallel investigation by Taylor et al. (submitted) using Gaussian process emulation techniques.

[2] We note that this relation connects the scale radius a to the total mass of the system. However, the massive elliptical galaxies that host the dominant PTA GW sources, are bulge dominated so that $M_{\rm bulge}$ can be taken as a fair proxy of the total stellar mass.

therefore be written as:

$$h_c^2(f) = \int dz \int d\mathscr{M} \frac{d^2n}{dz d\mathscr{M}} h_{c,\rm fit}^2\left(f\frac{f_{p,0}}{f_{p,t}}\right) \times \left(\frac{f_{p,t}}{f_{p,0}}\right)^{-4/3} \left(\frac{\mathscr{M}}{\mathscr{M}_0}\right)^{5/3} \left(\frac{1+z}{1+z_0}\right)^{-1/3} \quad (1)$$

where $h_{c,\rm fit}$ is an analytic fit to the spectrum produced by a reference binary at redshift z_0 with chirp mass $\mathscr{M}_0$ and a given eccentricity e_0 at an arbitrary decoupling frequency f_0. These two latter parameters define the peak frequency of the emitted GW spectrum $f_{p,0}$ for this reference binary. Equation (1) states that the overall GW spectrum from a given MBHB population can be generated from this reference $h_{c,\rm fit}$ via appropriate power-law scaling of the the chirp mass, redshift, decoupling frequency and eccentricity. Individual contributions must then be integrated over the MBHB mass function $d^2n/dzd\mathscr{M}$; the number of binary mergers per co-moving volume, redshift and (rest-frame) chirp mass interval. The integration limits of equation (1) are set to $0 \leq z \leq 5$ and $10^6 \leq \mathscr{M}/M_\odot \leq 10^{11}$, and following M16 we pick

$$\frac{d^2n}{dz d\log_{10}\mathscr{M}} = \dot{n}_0 \left[\left(\frac{\mathscr{M}}{10^7 M_\odot}\right)^{-\alpha} \exp^{-(\mathscr{M}/\mathscr{M}_*)}\right] \times \left[(1+z)^\beta \exp^{-(z/z_*)}\right] \frac{dt_R}{dz}, \quad (2)$$

where t_R is the time in the source rest-frame and dt_R/dz is given by the standard time-redshift cosmological relation (in this work we assume $H_0 = 70\rm km\ s^{-1}Mpc^{-1}$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $\Omega_k = 0$). The differential merger rate density of equation (2) is described by five parameters. $\dot{n}_0$ is the merger rate density normalization. β and z_* describe the redshift evolution of the rate. In particular, β controls the low-redshift power-law slope and z_* the high-redshift cut-off for the distribution; the peak of the merger rate corresponds to a redshift $(z_*\beta - 1)$. α and $\mathscr{M}_*$ are the free parameters of the Schechter function describing the mass distribution. In addition to those, the computation of the GWB in equation (1) requires the specification of the MBHB eccentricity e_t when they decouple from their environment and the evolution is dominated by GW emission[3], giving a total of six model parameters. Decoupling takes place when the condition that stellar scattering and GW emission extract energy from the MBHB at the same rate. This occurs at a frequency f_t, defined by (see PaperI)

$$f_t = 0.356\,\rm nHz \left(\frac{1}{F(e)}\frac{\rho_{i,100}}{\sigma_{200}}\right)^{3/10} \mathscr{M}_9^{-2/5}, \quad (3)$$

where the mass density of the stellar environment is $\rho_{i,100} = \rho_i/(100 M_\odot \rm pc^{-3})$, the velocity dispersion of the stars is the bulge is $\sigma_{200} = \sigma/(200\,\rm km\,s^{-1})$ and the MBHB total mass is $\mathscr{M}_9 = \mathscr{M}/(10^9 M_\odot)$. Expressions for $\rho_{i,100}$ and σ_{200} can be found in PaperI (equations 28 and 30). Note that ρ_i is a function of the inner slope of the adopted density profile. Here we adopt a Dehnen model with $\gamma = 1$, which results in shallow nuclear stellar density profiles that are typical of massive elliptical galaxies.

The characteristic amplitude described by equation (1) is a power-law with a low frequency turnover due to eccentricity and environmental effects. At high frequency, however, because of small number statistics, the actual signal is characterized by sparse resolvable systems outshining the overall GWB. Sesana et al. (2008) showed that the correct estimate of the unresolved GWB level can be recovered by setting an upper limit $\bar{\mathscr{M}}$ to the mass integral given by the condition

$$N_{\Delta f} = \int_{f-\Delta f/2}^{f+\Delta f/2} df \int_{\bar{\mathscr{M}}}^{\infty} d\mathscr{M} \int_0^\infty dz \frac{d^3N}{df dz d\mathscr{M}} = 1, \quad (4)$$

where $d^3N/(df dz d\mathscr{M})$ is the number of individual sources per unit chirp mass, redshift and frequency, which can be directly computed from $d^2n/dzd\mathscr{M}$ (see Sesana et al. 2008, for details), and the integral is performed over the frequency bin $\Delta f = 1/T$. The net effect is that the spectrum has a mass function dependent high frequency steepening, that can provide further information about the underlying MBHB population. Note that this is set solely by the MBHB mass function and does not introduce further parameters to the model. Examples of spectra highlighting both the low frequency turnover and the high frequency steepening are shown in figure 1.

The model was chosen to capture the expected qualitative features of the cosmic MBH merger rate without restricting to any particular merger history; for example, it can reproduce rates extracted from merger tree models (Volonteri et al. 2003; Sesana et al. 2008), and large scale cosmological simulations of structure formation (Springel et al. 2005; Sesana et al. 2009).

3 BACKGROUND DETECTION THEORY

The S/N ρ imprinted by stochastic GWB in a PTA can be written as (Moore et al. 2015; Rosado et al. 2015)

$$\rho^2 = 2 \sum_{i=1,N} \sum_{j>i} T_{ij} \int \frac{\Gamma_{ij}^2 S_h^2}{(S_n^2)_{ij}} df. \quad (5)$$

We now proceed to define and discuss all the elements appearing in equation (5). T_{ij} is the time span over which observations for pulsars i and j overlap. We will make from here on the simplifying assumptions that all pulsars are observed for the same timespan T (typically 10 years or more) and therefore $T_{ij} = T, \forall (i,j)$. However, we should bear in mind that this is generally not the case for real PTAs. The double sum runs over all the possible pairs of pulsars in the array and Γ_{ij} are the Hellings & Downs correlation coefficients (Hellings & Downs 1983)

$$\Gamma_{ij} = \frac{3}{2}\gamma_{ij}\ln(\gamma_{ij}) - \frac{1}{4}\gamma_{ij} + \frac{1}{2} + \frac{1}{2}\delta_{ij}, \quad (6)$$

where $\gamma_{ij} = [1 - \cos(\theta_{ij})]/2$, and θ_{ij} is the relative angle between pulsars i and j. S_h, S_n are the spectral densities of the signal and the noise respectively. The former is connected to the characteristic amplitude of the signal $h_c(f)$ given in equation (1) via:

$$S_h = \frac{h_c^2}{12\pi^2 f^3}, \quad (7)$$

where f is the considered frequency. The latter has to be handled with care, especially in the limit of a strong GWB signal. For a pulsar i characterized by random Gaussian irregularities described by a root mean square (rms) value σ_i^2, the power spectral density (PSD) of the noise is given by

$$P_i = 2\sigma_i^2 \Delta t, \quad (8)$$

where Δt is the interval between subsequent observations (typically a week to a month, in current PTAs). If red processes were not

[3] In this pilot study, we make the simplistic assumption that all MBHBs have the same eccentricity at decoupling. In general, MBHBs are expected to have a range of eccentricities when they decouple from their environment. Nonetheless, one can still try to model the population with a single parameter e_t, representing the typical MBHB eccentricity.

present in the data, one might then expect a PSD of the noise equal to P_i in the whole sensitivity window down to $1/T$. However, fitting for the spin first and second derivatives when constructing the pulsar timing model subtracts a quadratic function to the timing residual, effectively absorbing power at the lowest frequency bins, should a red signal be present.

To mimic the effect of the timing model we empirically write

$$P_i = 2\sigma_i^2 \Delta t + \frac{\delta}{f^5}, \tag{9}$$

where δ is a constant that depends on the parameters of the observations. We find that a good fit to the low frequency behaviour of the published EPTA, NANOGrav and PPTA sensitivity curves is provided by setting

$$\delta = 5\times 10^{-49}\left(\frac{10\mathrm{yr}}{T}\right)^5\left(\frac{\sigma_i}{100\mathrm{ns}}\right)^2\frac{\Delta t}{2\mathrm{weeks}}. \tag{10}$$

The scaling in equation (10) ensures that the curve maintains the same shape when varying the array parameters, reproducing the power absorption at the two lowest frequency bins (see Fig. 1). Moreover the PSD of the noise S_n is not only given by limitations in the pulsar stability, quadratic spindown fitting, and other sources of noise. The very same signal S_h contributes an equal amount to the noise as to the signal itself, because half of the GWB (the pulsar term) is uncorrelated. However, the smoking-gun of a GWB is provided by its distinctive quadrupole correlation described by the Γ_{ij} coefficients. Therefore only the correlated part of the signal (i.e. the Earth term) contributes to the construction of the detection statistic and to the build-up of the S/N. The pulsar term will just produce an uncorrelated common red noise in all pulsars with PSD S_h. Therefore the power spectral density of the noise has to be written as (Rosado et al. 2015):

$$S_{n,ij}^2 = P_iP_j + S_h[P_i+P_j] + S_h^2(1+\Gamma_{ij})^2. \tag{11}$$

Note that equation (11) reduces to $S_{n,ij}^2 = P_iP_j$ in the weak signal limit. Note, moreover, that this implies that it does not matter how strong the signal is, the integrand of equation (5) is at most of the order $\Gamma_{ij}^2 \ll 1$. This means that *only* with a large number N of pulsars is it possible to produce a confident detection of a GWB with an high ρ. This is easy to see if we make the simplifying assumptions that T, Δt and σ_i are the same for all pulsars. Moreover, we shall assume a sufficiently high number of randomly distributed pulsars in the sky, therefore substituting the individual Γ_{ij} with their average value $\Gamma = 1/(4\sqrt{3})$. Equation (5) can then be written as

$$\rho^2 = 2T\Gamma^2\int\frac{S_h^2}{S_n^2}\sum_{i=1,N}\sum_{j>i}df, \tag{12}$$

which reduces to

$$\rho^2 = T\Gamma^2N(N-1)\int\frac{S_h^2}{S_n^2}df. \tag{13}$$

In an actual observation, the GWB is resolved in bins $\Delta f = 1/T$. We can therefore divide the frequency domain in intervals $\Delta f_i = [i/T,(i+1)/T]$ centred at $f_i = (2i+1)/(2T)$ and compute the S/N in each individual frequency bin as

$$\rho_i^2 = T\Gamma^2N(N-1)\int_{\Delta f_i}\frac{S_h^2}{S_n^2}df \approx \Gamma^2N(N-1)\frac{S_h^2}{S_n^2} \tag{14}$$

The total S/N of the observation is then simply obtain by summing in quadrature over the frequency bins

$$\rho = \left(\sum_i\rho_i^2\right)^{1/2}. \tag{15}$$

Note that in the limit of $S_h \gg P$ in a given frequency bin, equation (16) reduces to

$$\rho_i^2 = \frac{\Gamma^2}{1+\Gamma^2}N(N-1). \tag{16}$$

Therefore, in the presence of a strong signal in M frequency bins, one gets an approximate S/N

$$\rho = \left(\frac{\Gamma^2}{1+\Gamma^2}MN(N-1)\right)^{1/2} \approx \Gamma NM^{1/2}. \tag{17}$$

Where we used the fact that $\Gamma \ll 1$ and $N \gg 1$. Equation (17) was obtained through a number of drastic simplifications, nonetheless it gives a sense of the maximum S/N one can obtain assuming a strong signal in an ideal array. Since $\Gamma \approx 0.14$, a total S/N≈ 5 in the lowest few frequency bins can only be achieved with approximately $N = 20$ equally good pulsars.

4 SIMULATING OBSERVATIONS

Once ρ_i has been computed at each frequency bin, we can then use the general fact that, if h is a signal described by an amplitude A, then $\rho = (h|h)$ and $\sigma_A^{-1} = (\partial h/\partial A, \partial h/\partial A)^{1/2} = (h/A, h/A)^{1/2}$. Therefore

$$\frac{\sigma_A}{A} = \sigma_{\ln A} = \frac{1}{\rho}. \tag{18}$$

To simulate observations, we therefore compute the S/N ρ_i at each frequency bin. If $\rho_i > 1$, we then assume an observed signal with amplitude $A_i = h_c(f_i)$ and error described by a log-normal distribution with width given by equation (18). Note that, by doing this we are ignoring any stochastic fluctuation in the measured amplitude of the signal. In reality, the error on the observation will be generally centred at $A_i \neq h_c(f_i)$, with a scatter of the order of the error on the measurement. We make this choice because our main aim is to investigate to what level the MBHB population model can be constrained *in principle*, independent of statistical variations inherent to the observations. If $\rho_i < 1$ then we assume no signal is detected in the frequency bin, and only an upper limit can be placed. To define what the upper limit is, we notice that, by means of equation (7), equation (16) can be written as a ratio of the characteristic signal and an equivalent characteristic noise, i.e.,

$$\rho_i = \frac{h_c^2}{h_n^2} \tag{19}$$

where,

$$h_n = [N(N-1)]^{1/4}\left(12\pi^2f^3\frac{S_n}{\Gamma}\right)^{1/2}. \tag{20}$$

Therefore, when $\rho_i < 1$ we place a 68% (1σ) upper limit at $h_{n,i}$, calculated at the central frequency f_i of the bin.

Examples of signal generation are shown in Fig. 1 for spectra with $A = 10^{-15}$ at $f = 1/1$yr and an array with $N = 20$, $\sigma = 100$ns, $T = 15$yr, $\Delta t = 1$ week. This setup results in a detection with moderate S/N, $\rho \approx 5$, and with $\rho_i \approx 2$ in the few lowest frequency bins. The equivalent h_n of equation (20) is depicted as a black solid line. Note, however that for clarity of representation, we ignored here the contribution of S_h to the noise (when that is taken into account, $h_n = h_c$ whenever $\rho_i > 1$). Note also that, despite the large h_c difference of the two signals, the difference in S/N between them is only about 20%. This is because, as stressed above, in the strong signal limit the S/N of the signal is limited by the GWB uncorrelated self-noise.

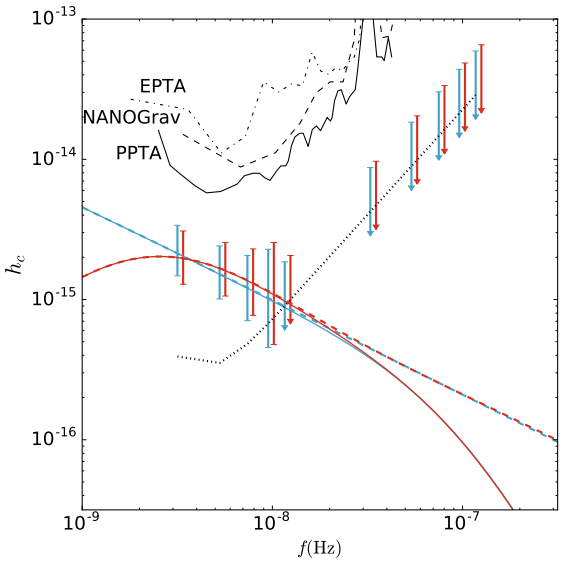


Figure 1. Examples of simulated detections for two different spectral shapes. Signal models correspond to the default MBHB population with parameters defined in Section 4.1 and high eccentricity ($e_t = 0.9$, red) and almost circular ($e_t = 0.01$), blue). For each model, solid lines are the theoretical spectra including the high frequency steepening due to the mass upper limit defined by equation (4), dashed lines depict spectra excluding this feature (therefore with $h_c \propto f^{-2/3}$ at high frequency) for comparison. Error bars centred around the model value are the observed amplitudes with associated uncertainties when $\rho_i > 1$, and downward arrows represent upper limits equal to $2h_n$ (i.e. 2σ) when $\rho_i < 1$ at their base. The black dotted line is the characteristic noise level h_n *excluding* the contribution of the GW signal to the noise budget. Black lines in the upper part of the figure are current EPTA, NANOGrav and PPTA limits. We assume 15 years of observation of 20 pulsars at 100ns rms.

4.1 Simulation setup

To setup a specific simulation, one has to define both the properties of the GWB (i.e. the six parameters $\dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t$ defining the MBHB population) and of the PTA employed for detection (i.e. the four parameters $N, \sigma, T, \Delta t$ defining the sensitivity of the array).

Unless otherwise stated, we use a MBHB mass function defined by $\dot{n}_0 = 10^{-4}\mathrm{Mpc}^{-3}\mathrm{Gyr}^{-1}, \beta = 2, z_* = 2, \alpha = 0, \mathscr{M}_* = 10^8 M_\odot$. The normalization $\dot{n}_0$ and the redshift dependence β are chosen to be consistent with current estimates of the galaxy merger rate (Lin et al. 2004; de Ravel et al. 2009; Lotz et al. 2011). α and $\mathscr{M}_*$ are chosen to ensure that the shape of the MBHB mass function is consistent with that of nuclear MBHs as inferred from direct measurements and MBH-galaxy scaling relations (e.g. Shankar et al. 2004; Hopkins et al. 2007). The adopted parameters result in a GWB with characteristic strain at $f = 1\mathrm{yr}^{-1}$ of $A \approx 5 \times 10^{-16}$, fully consistent with current upper limits. We explore different eccentricities at decoupling and we report results for the illustrative cases of quasi circular and highly eccentric binaries, defined by $e_t = 0.01$ and $e_t = 0.9$ respectively.

We make the simplifying assumption that all pulsars are observed for the same timespan T, with the same cadence Δt and have the same rms σ. Note that our main results are nevertheless general, since these assumptions only affect the computation of the S/N and do not enter in the subsequent analysis of the GWB spectral shape. We consider four different array scenarios:

(i) case *PPTA15*: in this case we simply use the curve provided by (Shannon et al. 2015), which is representative of current PTA capabilities and results in an upper limit of $A = 10^{-15}$.

(ii) case *IPTA30*: $N = 20$, $\sigma = 100$ns, $T = 30$yr, $\Delta t = 1$ week. This PTA results in a detection S/N≈ 5 and is based on a future extrapolation of the current IPTA, without the addition of new telescopes.

(iii) case *SKA20*: $N = 100$, $\sigma = 50$ns, $T = 20$yr, $\Delta t = 1$ week. This PTA results in a high significance detection with S/N$\approx 30-40$, which will be technically possible in the SKA era.

(iv) case *ideal*: $N = 500$, $\sigma < 1$ns, $T = 30$yr, $\Delta t = 1$ week. This is likely beyond SKA capabilities but provides useful insights of what might be achievable in principle.

As stated above, for each simulations we compute the the S/N ρ_i at each frequency bin. If $\rho_i > 1$, we then assume an observed signal with amplitude $A_i = h_c(f_i)$ and error described by a log-normal distribution with width given by equation (18). If $\rho_i < 1$ then we place an upper limit at h_n as defined by equation (20).

5 DATA ANALYSIS METHOD

As in M16, our aim is to constrain the astrophysical population of merging MBHB given some PTA data. The data consists of an array of measurements and upper limits on the GW spectrum at different frequency bins, as described in the previous section. In M16, we assumed circular binaries and an $f^{-2/3}$ power law for the spectrum, meaning that all the information from the background could be summarised with two numbers, an upper limit or detection with some confidence at a given frequency, which we chose to be one over one year. In this paper, we allow for eccentric binaries evolving via scattering of background stars and a finite number of sources at high frequencies, both of which result in a spectrum that is different from the $f^{-2/3}$ power law. Therefore, the shape of the spectrum over the frequency band encodes much more information. In this section, we describe our strategy to infer the astrophysical properties of the merging MBHB population from PTA measurements.

We denote our astrophysical model (section 2) as M and our data (section 4.1) as d. Our intention is to infer the model parameters θ, given a specific measurement. We start from Bayes theorem,

$$p(\theta|d,M) = \frac{p(\theta|M)p(d|\theta,M)}{p(d|M)}, \qquad (21)$$

where $p(\theta|d,M)$ is the posterior distribution for the model parameters given the data and the model, $p(\theta|M)$ is the prior, representing any initial knowledge we have on the parameters given the specific model, $p(d|\theta,M)$ is the likelihood for the data given the model and some values of the parameters and finally $p(d|M)$ is the evidence.

As described in section 2, our model has six parameters $\theta = \dot{n}_0, \beta, z_*, \alpha, \mathscr{M}_*, e_t$. Unless otherwise stated, for our analysis we choose priors as follows: the parameters β, z_*, α and e_t are all uniformly distributed in the ranges $\beta \in [-2,7]$, $z_* \in [0.2,5]$, $\alpha \in [-3,3]$, $\log_{10}\mathscr{M}_*/M_\odot \in [6,11]$, and $e_t \in [10^{-6}, 0.999]$. The prior for the merger rate parameter, $\dot{n}_0$ is log-uniform for $\dot{n}_0 \in [10^{-20}, 10^3]$ and uniform in $\dot{n}_0$ for $\dot{n}_0 < 0$, thus allowing for the possibility of no mergers. We note that although specific combinations of parameters can mimic MBHB merger rates extracted from semi-analytic merger tree models (Sesana et al. 2008), cosmological simulations of galaxy formation (Sesana et al. 2009; Kelley et al. 2016) and observations of galaxy pairs (Sesana 2013b), the adopted prior range is highly uninformative and allows for exotic MBHBs mass functions that are not necessarily related to galaxy mergers. For example, the upper

limit in $\dot{n}_0$ is solely dictated by the constraint that all the dark matter in the Universe is formed by merging MBHs.

The functional form of the likelihood function we adopt depends upon the type of data in each frequency bin. For a given spectrum, there are two possible observational outcomes in a specific frequency bin; either a GWB detection at $A_{\rm det}(f)$, or a non-detection, resulting in an upper limit based on the PTA sensitivity at that frequency $A_{\rm ul}(f)$. For an upper limit, we model the likelihood as a Fermi-like distribution,

$$p_{\rm ul}(d|A_{\rm trial}(f)) \propto \left\{ \exp\left[\frac{A_{\rm trial}(f) - A_{\rm ul}(f)}{\sigma_{\rm ul}(f)} \right] + 1 \right\}^{-1}, \qquad (22)$$

where $A_{\rm ul}$ is the measurement of the upper limit at frequency f, $\sigma_{\rm ul}$ controls the width of the tail-off and $A_{\rm trial}(f)$ is the GWB given by our model for a set of parameters drawn from the prior. The value of σ_{ul} controls the steepness of the function as it transits from 0 at $A_{\rm trial} \gg A_{\rm ul}$ to some constant value for $A_{\rm trial} \ll A_{\rm ul}$ and can be adjusted so that, for example $p(A_{\rm trial} < A_{\rm ul}) = 68\%$. In our simulations, $A_{\rm ul} = h_n$ as described in section 4. We are therefore using the sensitivity of the PTA as a proxy for the 68% (or *1-sigma*) upper limit when the signal is not detected.

In the case of a detection, for the likelihood we choose a gaussian in the logarithm of the GWB measurement $A_{\rm det}$,

$$p_{\rm det}(d|A_{\rm trial}(f)) \propto \exp\left\{ -\frac{[\log_{10} A_{\rm trial}(f) - \log_{10} A_{\rm det}(f)]^2}{2\sigma_{\rm det}(f)^2} \right\}, \qquad (23)$$

where $\sigma_{\rm det}$ is the error on the detection measurement as described in section 4 and $A_{\rm trial}$ is again the value of the GWB given by parameters sampled by the prior. As the dataset d consists of a collection of GWB measurements across the frequency spectrum, we need to combine the likelihood of all the frequency bins in our data. We assume statistical independence among the various frequency bins and thus compute the overall likelihood by multiplication of the likelihoods (either an upper limit or a detection) from each bin. Note that, when we combine bins with detections to bins with upper limits, we consider the lowest frequency upper limit and five further points spaced by ten bins. This is because bins become much denser at high frequency and considering all the upper limits slows done the likelihood computation substantially. We checked that this does not affect our results, since the only constraining upper limit is always the one at the lowest frequency.

We explore the parameter space by means of a Nested Sampling algorithm (Skilling 2004). We use a tailored version of the parallel implementation of Nested Sampling given in Del Pozzo & Veitch (2015) which is similar in spirit to Veitch & Vecchio (2010) and Veitch et al. (2015). For all the analysis presented in this work we set the number of live points to be $N \sim 2{,}000$ owing an average number of posterior samples $\sim 5{,}000$.

6 RESULTS AND DISCUSSION

In this section we present and discuss in detail the results of our simulations. We will start with the interpretation of upper limits and then move to the case of detections with small and large S/N. We stress that, unless otherwise stated, astrophysical interpretation is constructed uniquely on the basis of PTA observations, i.e. we do not use any additional constraints on the MBHB population (besides the wide, non-informative prior range of the model parameters). PTA inference can prove significantly more constraining if combined with independent information. For example one can assume a narrow prior on the MBHB merger rate and mass function based on simulations or observations of merging galaxies (e.g. Sesana 2013b). However, we caution that such information is often *indirect* and requires theoretical modelling subject to several assumptions.

6.1 Upper limits

We first consider the case of an upper limit and we take as example the most stringent constraint imposed by the PPTA of $A < 10^{-15}$ at $f = 1{\rm yr}^{-1}$. Although PTAs often quote limits at $f = 1{\rm yr}^{-1}$, those are the result of the integrated array sensitivity across the relevant frequency band. This is shown in the upper-left panel of Fig. 2; according to the analysis framework developed in section 2, we assume at each frequency bin a 95% upper-limit given by the dashed curve and run our analysis. Consistent with M16, the results shown in Fig. 2 indicate that current PTA upper limits alone return little astrophysical information, and only loose upper bounds can be placed on the MBHB mass function (upper-right panel) and redshift (lower-left panel) distribution. Those are defined by integrating equation (2) in the redshift range [0,5] and in the mass range $[10^6 M_\odot, 10^{11} M_\odot]$, respectively. The triangle plot in the lower-left panel shows that the posterior distributions of the model parameters are essentially flat (β and z_* are not shown, as they are always flat due to strong degeneracy with $\dot{n}_0$), with the exception of $\dot{n}_0$, which is found to be $< 2.5 \times 10^{-3} {\rm Mpc}^{-3} {\rm Gyr}^{-1}$ at the 95% level. This constraint becomes interesting when compared to independent information on galaxy merger rates. Several observational studies place the merger rate density of massive galaxies at $z < 1$ to be around few$\times 10^{-4} {\rm Mpc}^{-3} {\rm Gyr}^{-1}$ (Lin et al. 2004; Lotz et al. 2011; Xu et al. 2012). In fact, this is in essence the reason why some tension between PTA upper limits and vanilla MBHB assembly models was highlighted by Shannon et al. (2015). We will return in more depth on this point in a companion paper (Middleton et al. in preparation). A tighter upper limit, constraining $\dot{n}_0$ to be less than $10^{-5} {\rm Mpc}^{-3} {\rm Gyr}^{-1}$ might rule out a naive one-to-one correspondence between galaxy and MBHB mergers, indicating that delays, stalling or high MBHB eccentricities play a major role in the dynamics.

6.2 PTA detection constraints on model parameters

We turn now to the implication of a future PTA detection. We discuss two distinct MBHB populations corresponding to our default mass function model (with parameters given in section 4.1) and defined by decoupling eccentricity $e_t = 0.01$ (circular case) and $e_t = 0.9$ (eccentric case).

6.2.1 *Circular case*

Results for the circular case are shown in Fig. 3 to which we refer in the following discussion. In the *IPTA30* scenario (left column), the signal is detected in the lowest eight frequency bins, with total S/N≈ 6. At $f < 10$ nHz the spectrum is well constrained (upper panel), and the reconstructed MBHB mass function and redshift distribution (central panels) are consistent with the injected values. Note, however, that astrophysical constraints are quite poor; even around $\mathscr{M} = 3 \times 10^8 M_\odot$, where the mass function is best constrained, the 68% confidence interval spans about two order of magnitude, and so does the high mass cut-off. The triangle plot in the lower panel provides more insight into the reconstruction of the model parameters. In general, the posteriors of all the parameters

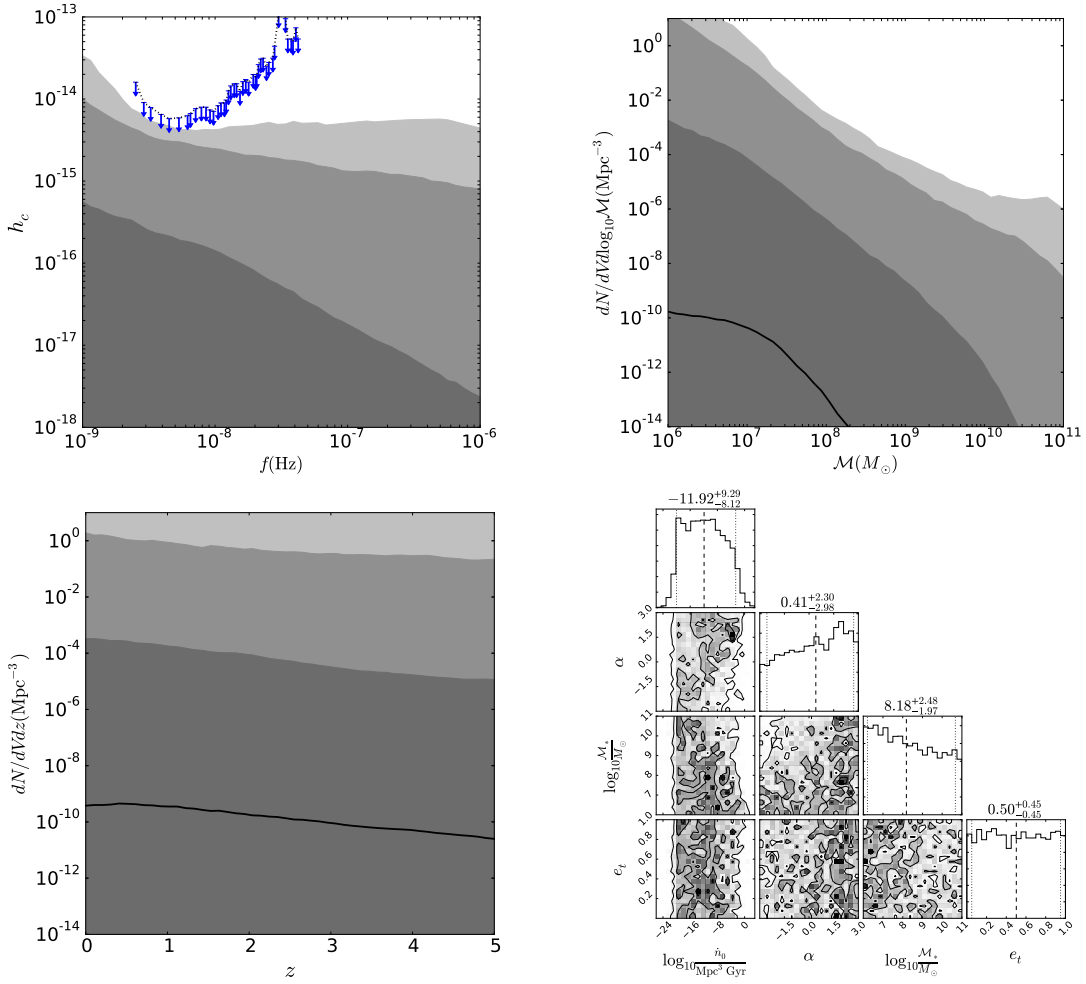


Figure 2. Implication of a 95% upper-limit of $A(f = \mathrm{yr}^{-1}) = 1 \times 10^{-15}$, which corresponds to the most stringent PTA upper limit to date. The posterior for the spectrum (top left), mass (top right) and redshift functions (bottom left) are shown as shaded areas, with the 68%, 95% and 99.7% confidence regions indicated by progressively lighter shades of grey, and the solid black line marking the median of the posterior. The dotted line with downwards pointing arrows in the top left panel is the 95% upper limit from Shannon et al. (2015). The bottom right triangular plot shows the two-dimensional posteriors for each model parameter pairs, together with their one dimensional marginalised distributions. The lines in each one dimensional distribution mark the median (dashed) and the central 90% (dotted) of the posterior, with the numerical values indicated above each plot.

are consistent with the injected values, however the distributions are fairly broad and the contour plots unveil several correlations among model parameters, the most important of which will be investigated later on.

The situation quantitatively improves, but is qualitatively unaltered, in the *SKA20* scenario, shown in the right column. Here the signal is detected in 13 frequency bins, with a total S/N$\approx$ 35. The h_c spectrum is extremely well reconstructed up to 20nHz and the median of the recovered mass and redshift functions match the injected ones almost exactly (central panel); uncertainties are still large though, and the posterior distributions of the model parameters improve only marginally. The characteristic mass scale $\mathscr{M}_*$ is slightly better constrained and, compared to the *IPTA30* case, there is a stronger preference for circular binaries, although higher eccentricity cannot be ruled out.

6.2.2 *Eccentric case: parameter degeneracies*

The eccentric case is shown in Fig. 4. Again, in the *IPTA30* scenario (left column panels) the signal is detected in the nine lowest frequency bins, with total S/N$\approx$ 5. The recovered GW spectrum is consistent with the injected one, but errors are large and the shape can be hardly determined. The triangle plot in the lower-left panel shows that it is difficult to recover model parameters. Posteriors are consistent with injected values, but the distributions are hardly informative.

Moving to the *SKA20* case (right column panels), we see a clear improvement on the reconstruction of the spectrum (upper panel), but the preferred mass function appears quite offset with respect to the original injection (second panel from the top). Posterior distributions in the triangle plot (lower panel) are now more informative and reveal more defined degeneracies. Particularly interesting is the $\int$-shaped posterior in the $e_t - \mathscr{M}_*$ panel (already visible in the *IPTA30* case). The degeneracy stems from the mass dependence

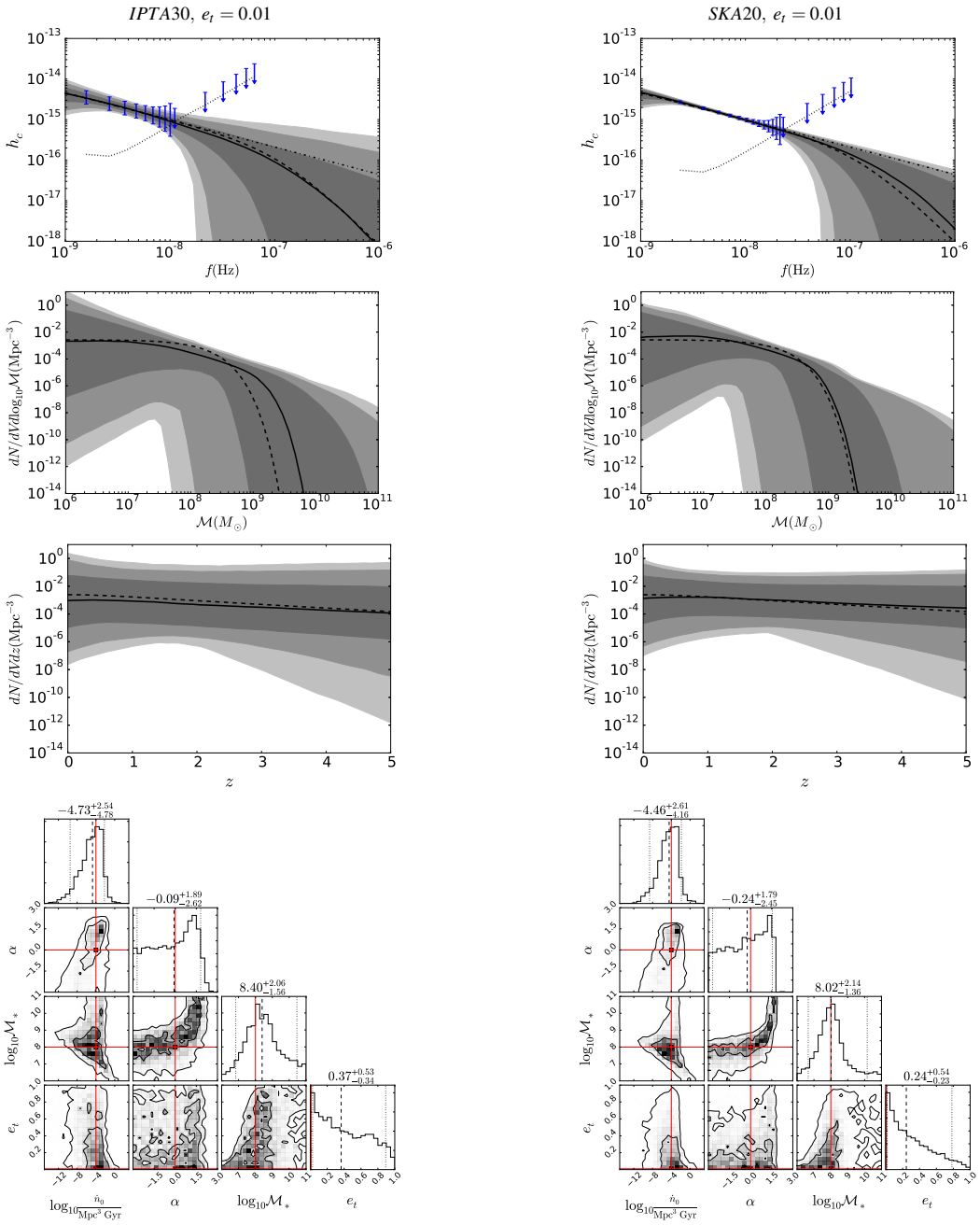


Figure 3. Implication of a PTA detection at a moderate (S/N≈5, left column) and high (S/N≈35, right column) significance, assuming a MBHB population with default mass function parameters and almost circular ($e_t = 0.01$) eccentricity at decoupling. As in Fig. 2, the posterior for the spectrum, mass and redshift functions (in descending order from the top) are shown as shaded areas, with the 68%, 95% and 99.7% confidence regions indicated by progressively lighter shades of grey, and the solid black line marking the median of the posterior. In each of those panels, the dashed black line indicates the injected model. In the top panels the vertical blue bands indicate the 68% confidence interval of the observed signal amplitude at each frequency bin, and the downward pointing arrows at higher frequency mark the 95% upper limits. The dotted line is the nominal 1σ sensitivity of the considered PTA, as defined by equation 20, where the contribution of S_h to the noise has been omitted (see section 4 for details). The dot-dash black line shows the simulated spectrum assuming no drop in high mass sources. The lower triangular plots show the two-dimensional posteriors for each model parameter pairs, together with their marginalised distributions. The injected parameter values are marked by red solid lines and the black lines in each one dimensional distribution mark the median (dashed) and the central 90% (dotted) of the posterior, along with the numerical values above each plot.

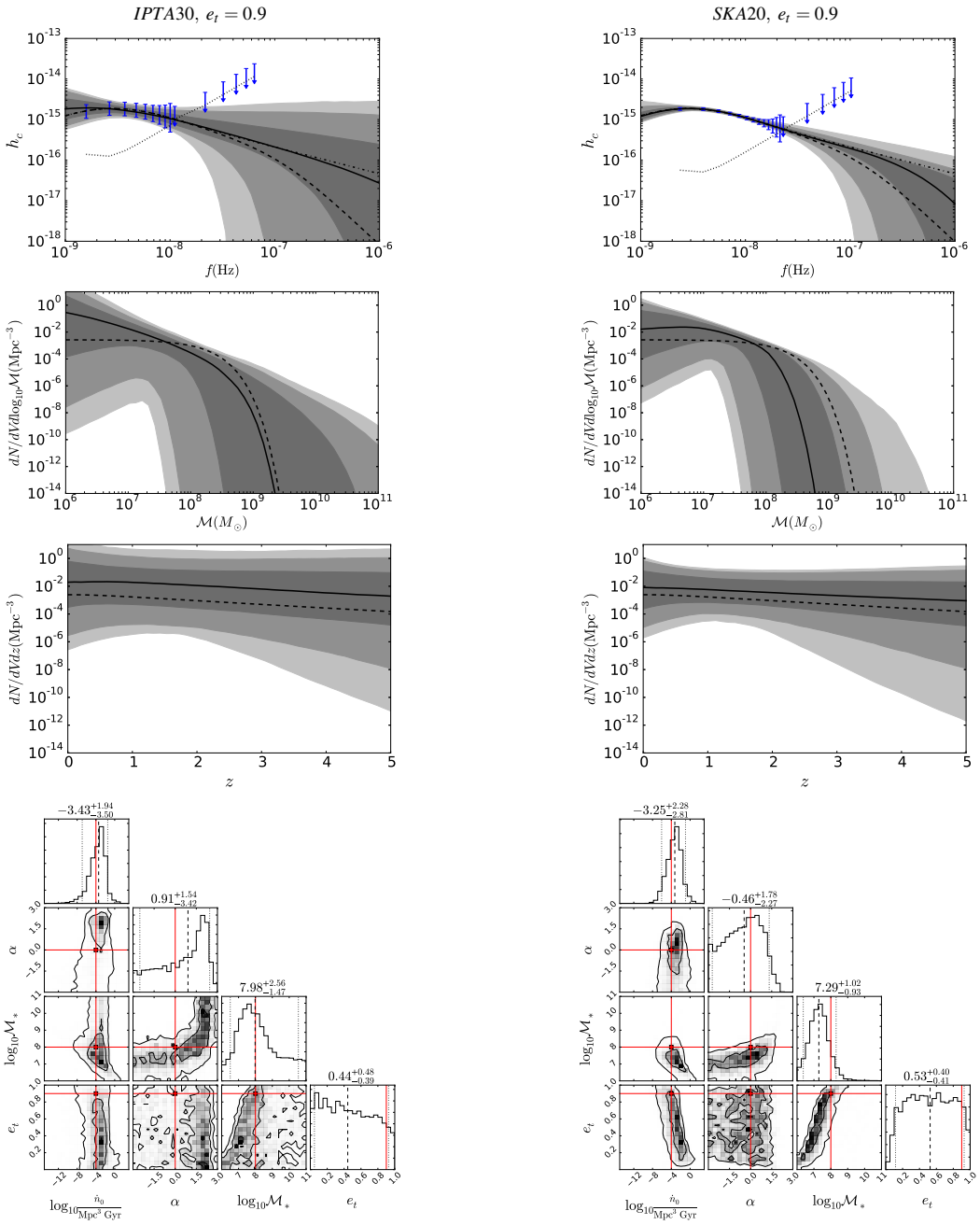


Figure 4. Same as figure 3 but assuming decoupling eccentricity of $e_t = 0.9$.

of the decoupling frequency in equation (3), i.e. from the fact that more massive MBHBs decouple at lower frequencies than lighter ones. In fact, for a given eccentric MBHB, the peak of the GW spectrum occurs at a frequency $f_p = \mathscr{F}(e_t) f_d$ (see equation 13 in PaperI), where $\mathscr{F}(e_t)$ is a monotonically increasing function of e_t. This means that, if we observe a turnover in the GWB at a given $\bar{f}$, there is an ambiguity in the determination of the decoupling eccentricity of the MBHB population. The signal can be dominated by lighter MBHB decoupling at higher f_d with lower e_t, or by heavier MBHB decoupling at lower f_d with higher e_t, giving rise to the $\int$-shaped contour in the $e_t - \log_{10}\mathscr{M}_*$ plane. Lighter black holes require a higher $\dot{n}_0$ to produce the observed signal level, however this is still well within the assumed prior. In practice, the detection of a turnover in the GWB, guarantees that MBHBs have some eccentricity at decoupling (which in our models always occur below the observable PTA frequency window), however cannot inform us on the value of their eccentricity, unless independent information on the MBHB mass function becomes available. This causes the peculiar shape of the e_t posterior seen in the lower-right panel of Fig. 4, in which the posterior is quite flat down to $e_t \approx 0.1$ and has a sharp decline disfavouring circular binaries.

6.3 Breaking degeneracies: the importance of detection at high frequencies

We saw in the previous section that parameter degeneracies prevent a precise characterization of the properties of the underlying MBHB population. This is because the GWB spectrum does not present sufficient structure to allow proper parameter estimation. In principle, the high frequency steepening of the GWB offers a tantalizing possibility of an independent measurement of the mass function parameters. In practice, unfortunately, the steepening generally occurs at $f > 30$ nHz where PTA sensitivity drops significantly. A measurement might be possible for MBHB population featuring a heavy-biased MBHB mass function, for which the steepening occurs already around $f \approx 10$ nHz. However, even in this case, errorbars on the detected amplitude at the highest frequency bins would be quite large, making a proper measurement of the drop problematic.

Although this is likely out of reach for current and planned PTA efforts, as a proof of principle we show here what information can be recovered with a measurement of the GWB spectrum up to $f = 5 \times 10^{-7}$ Hz, possible with our *ideal* array. Performing a parameter space exploration would be impractical, because for 30 years of observation, the signal would be observed in about 500 frequency bins, making the evaluation of the likelihood function prohibitively time consuming. We therefore interpolate the observations (with relative errorbars) in 20 equally log-spaced bins in the range $10^{-9} - 5 \times 10^{-7}$ Hz. Note that the total S/N of such detection is not much higher than the *SKA20*, however we will see that the high frequency extension makes a critical difference in the recovery of the MBHB population parameters (even if we are not using all the information enclosed in the original 500 frequency bins). This is shown in Fig. 5 for our standard MBHB population with $e_t = 0.01$ (left column) and $e_t = 0.9$ (right column). The upper panels show that, contrary to all previous cases, the high frequency steepening is now well characterized; this is the key element, because its shape depends on the MBHB mass function. Posterior distributions of the population parameters are shown in the lower triangle plots. The parameters defining the MBHB mass function are now well constrained and peak around the injected values; the cut-off mass scale $\mathscr{M}_*$ is determined within a factor of three and the slope α within ≈0.2. The recovery of the eccentricity is also much cleaner. Posteriors are still broad, but in the circular case one can confidently say that the typical eccentricity of the MBHBs is < 0.16 (95% confidence) although the posterior peaks at $e_t \approx 0.1$. This is because a non detection of a low frequency turnover is still consistent with mildly eccentric binaries at decoupling, even if the mass function parameters are fairly well determined. Similarly, for the eccentric case, one can state with 95% confidence that the typical eccentricity of the MBHBs is > 0.7 and the posterior is quite flat in the range $0.75 < e_t < 0.95$. One last thing to notice is that, even though the GWB spectrum is pinned down essentially exactly, there remains a remarkable uncertainty in the determination of the overall merger rate density $\dot{n}_0$. This is because of its intrinsic (not shown) degeneracy with the β and z_* parameters defining the redshift distribution of mergers. A low $\dot{n}_0$ normalization with a steep, positive redshift dependence β can result in the same GWB as a much higher $\dot{n}_0$ normalization with a flatter redshift dependence. Unless external information (see below) about the redshift evolution of the merger rate density is available, this degeneracy is unlikely to be disentangled on the basis of GWB measurements alone.

6.4 Adding independent constraints

So far, we considered what astrophysical information can be extracted by PTA observation only, deliberately ignoring any constraints on the MBHB population imposed by other observations. The motivation behind this agnostic choice is that those constraints are inevitably indirect, and involve either the rate of merging *galaxies* (e.g. Lin et al. 2004; Lotz et al. 2011; Xu et al. 2012) or the determination of the mass function of *single MBHs* (see for example Shankar et al. 2004). The conversion of a galaxy merger rate into a MBHB merger rate implies a number of uncertain assumptions about the relation between galaxy hosts and MBHs (Kormendy & Ho 2013, and references therein) and the effectiveness of the MBHB coalescence following galaxy mergers (e.g. McWilliams et al. 2014; Kelley et al. 2016); on the other hand, the mass function of individual MBHs in galaxy centers does not provide direct information on the properties of *merging* MBHBs.

It is nevertheless instructive and interesting to fold those indirect constraints into the analysis to understand to what extent PTA observation can improve the current state of the art of MBHB knowledge. Sesana (2013b) constructed a compilation of observationally-based MBHB merger distributions encompassing a wide uncertainty range in the galaxy merger rate and galaxy host-MBH relations. The outcome of the procedure is a loosely constrained MBHB mass function and redshift distribution resulting in a predicted GWB spanning almost two order of magnitudes in amplitude (at 99.7% confidence). In general, in the best constrained areas (chirp masses in the range $10^7 M_\odot$-$10^{8.5} M_\odot$ and $z < 1.5$), the uncertainty range spans about two orders of magnitudes. To incorporate this information in our analysis, we draw a large sample of populations from our unrestricted parameter range, and we accept only those for which the mass and redshift functions fall within the range constrained by the Sesana (2013b) models to update our prior. The restricted MBHB mass and redshift functions resulting from this procedure are shown as dotted areas in the central panels of Fig. 6. The restricted marginalized priors on the model parameters are shown in the triangular plots in the bottom panels and their median values and 90% confidence intervals are listed in table 1. Furthermore, since the merger rates do not constrain the MBHB eccentricity distribution at decoupling, we assume a flat prior on e_t. The resulting prior GWB spectrum is shown in the upper panels of Fig. 6. As expected the range of h_c is consistent with what is shown in figure 2 of Sesana (2013b). The difference in shape is due to the inclusion of the high frequency drop, and to the fact that we allow for very eccentric MBHB population, that cause a widening of the allowed h_c range at the low frequency. We assume that the true underlying MBHB population is described by our default models (shown with dashed lines), that falls well within the restricted prior range, and that MBHBs have $e_t = 0.9$ at decoupling.

The results of the analysis for two different PTAs are shown in Fig. 6 and measured parameter values are also listed in table 1. PTA observations in the foreseeable future (*IPTA30* case, left column) will place significant constraints to the higher end of the mass function, reducing the uncertainty range by more than one order of magnitude at $\mathscr{M} > 10^8 M_\odot$. The redshift function is poorly constrained, because the mass integral of the merger rate is dominated by the abundance of MBHBs with $\mathscr{M} < 10^8 M_\odot$, which remains poorly determined. This is also confirmed by the marginalised posterior distributions in the model parameters shown in the bottom panel. The posteriors on the overall merger rate $\dot{n}_0$ and on the redshift parameters β and z_* are essentially unaltered when compared to the prior, conversely, the prior knowledge of $\mathscr{M}_*$ is significantly

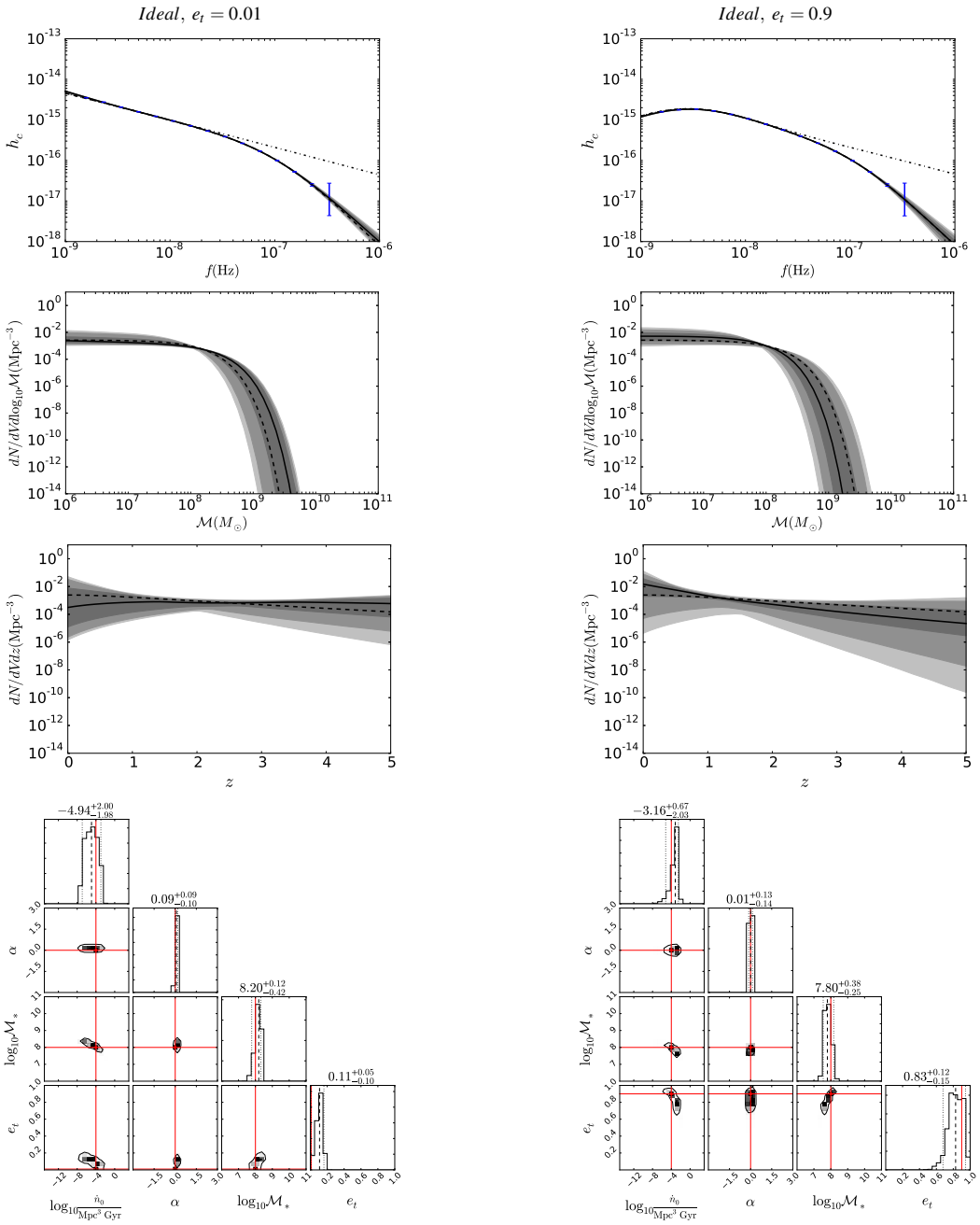


Figure 5. Implication of an ideal detection with 500 MSPs timed at sub-ns precision for 30 years. The injected model has default parameters with $e_t = 0.01$ (left column) and $e_t = 0.9$ (right column). Panel sequence and style as in Fig. 3.

updated with a 90% confidence interval shrinking by an order of magnitude. Note that, since $\mathscr{M}_*$ is decently constrained, the detection of the low frequency turnover is now quite informative, and eccentric binaries are favoured, with a posterior probability distribution correctly peaking around $e_t = 0.9$. In the ideal case, shown in the right column, the mass function is constrained almost exactly, and also our knowledge of the redshift evolution of the merger rate is significantly updated. The posterior distributions of the model parameters show that $\alpha, \mathscr{M}_*$ and e_t are pinned down with high accuracy. Moreover, also the degeneracy between the rate normalization and the redshift evolution is partially broken. The 90% credible interval on $\dot{n}_0$ shrinks by a factor of three compared to the prior, and the slope of redshift dependence β can be fairly well constrained, with a posterior peaking close to the injected value. This latter measurement is particularly interesting, because it would allow a direct comparison to the galaxy merger rate that is often observationally parametrised as being proportional to $(1+z)^\beta$.

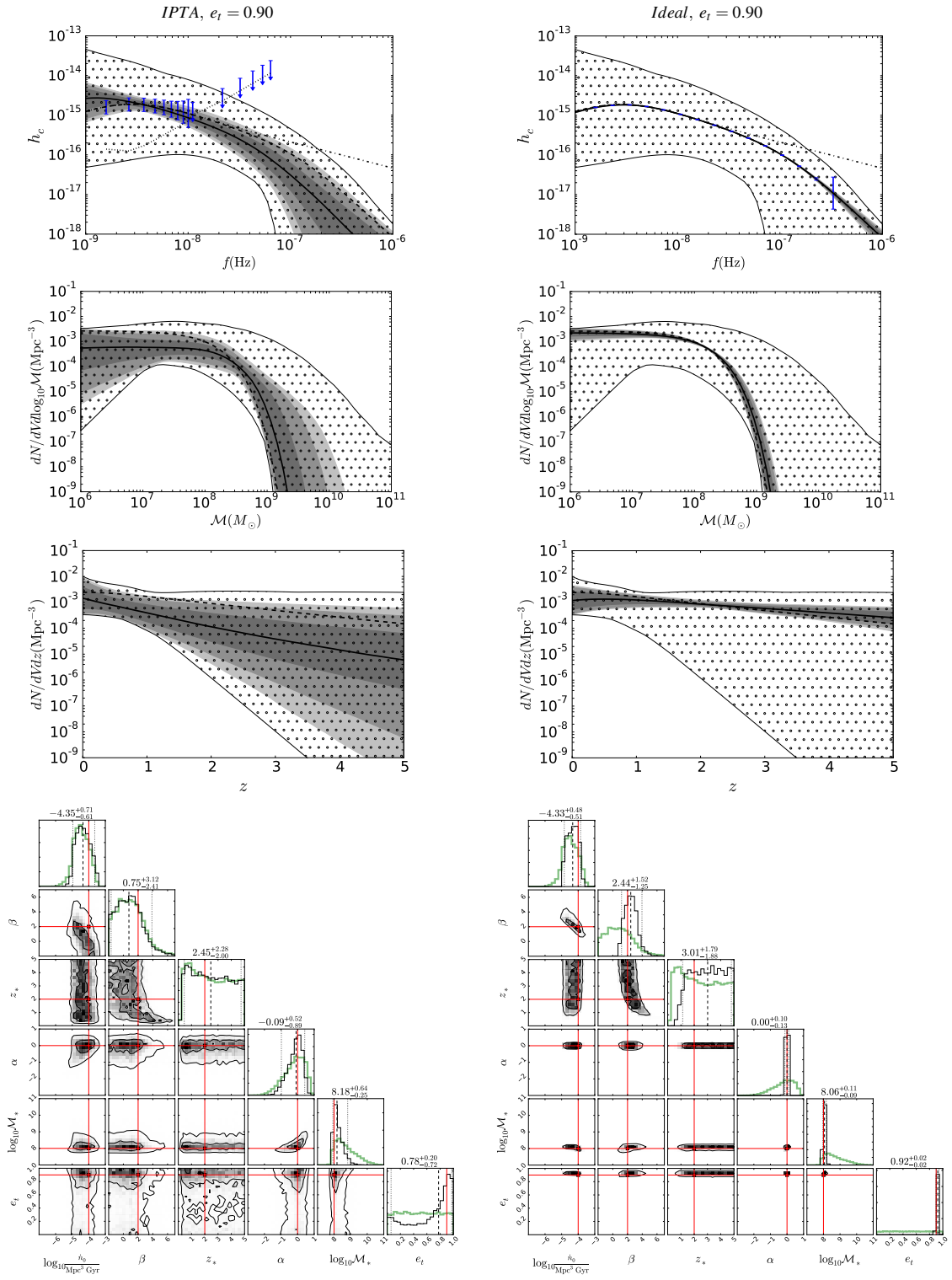


Figure 6. Effects of imposing external constraints on the MBHB mass and redshift distribution on the science return of PTA observations. The injected model has default mass function parameters and $e_t = 0.9$. In the left column we consider a moderate S/N detection with the *IPTA30* array, whereas the right panel is for an ideal detection as reported in Fig. 5. Panel sequence and style are as in Fig. 3. The additional dotted areas represent the restricted prior based on the astrophysical models of Sesana (2013b). The thick green histograms in the bottom panels show the marginalised prior distribution on the model parameters once the astrophysical constraint is imposed (see main text for full details).

parameter	prior	*IPTA30*	*ideal*
$\log_{10}\dot{n}$	$-4.47^{+0.73}_{-0.70}$	$-4.35^{+0.71}_{-0.61}$	$-4.43^{+0.48}_{-0.51}$
β	$0.81^{+3.29}_{-2.43}$	$0.75^{+3.12}_{-2.41}$	$2.44^{+1.52}_{-1.25}$
z_*	$2.39^{+2.36}_{-1.95}$	$2.45^{+2.28}_{-2.00}$	$3.01^{+1.79}_{-1.88}$
α	$-0.11^{+0.75}_{-1.25}$	$-0.09^{+0.52}_{-0.89}$	$0.00^{+0.10}_{-0.13}$
$\log_{10}\mathscr{M}_*$	$8.58^{+1.25}_{-0.65}$	$8.18^{+0.64}_{-0.25}$	$8.06^{+0.11}_{-0.09}$
e_t	$0.50^{+0.45}_{-0.45}$	$0.78^{+0.20}_{-0.72}$	$0.92^{+0.02}_{-0.02}$

Table 1. List of model parameters credible intervals for our constrained models. Each entry reports the median value together with the errors bracketing the 90% confidence regions. The three columns list the values defined by our restricted prior, and the posterior values as measured by the arrays *IPTA30* and *ideal*.

7 CONCLUSIONS AND OUTLOOK

We have performed the first extended investigation of the inverse problem for PTA data analysis, namely: given a PTA observation or upper limit, what constraints can be placed on the astrophysical properties of the underlying MBHB population? Our work expands on M16, by considering future detections in a sizeable frequency range, allowing us to fold into the analysis the information carried by the observed spectral shape of the GWB. To do so, we employed the semi-analytical model of Chen et al. (2016) that describes the MBHB population model with six physical parameters: five parameters shaping the redshift dependent mass function and an additional eccentricity parameter e_t that encapsulates the main effect of the MBHB coupling with the stellar environment. Depending on those parameters, the resulting GWB spectrum might show a significant departure from the nominal $f^{-2/3}$ power-law both at high frequency, due to small number statistics of the systems contributing to the signal (Sesana et al. 2008), and at low frequency, because of high eccentricity caused by interaction with stars in the inspiral phase. We explored to what extent such spectral features are recognizable and can be exploited to extract information from PTA observations. We assumed uninformative prior ranges in all the model parameters, consistent with the current absence of any secure direct observation of sub-parsec MBHBs emitting in the PTA relevant range.

We first used our analysis framework to assess the impact of current PTA upper limits, recovering the results of M16. In essence, a non detection can only impose a upper-bound on the overall merger rate density of MBHBs. Current PTA limits set this upper-bound to $\dot{n}_0 < 2.5 \times 10^{-3}$ (95% confidence), which is close to the range of currently measured galaxy merger rate densities, indicating that PTA observation are getting into the interesting astrophysical range. We then extended our investigation to a number of future detection scenarios: an IPTA-like array (*IPTA30*), and SKA-like array (*SKA20*) and an ideal array with 500 pulsars at sub-ns precision (*ideal*). In all cases, a GWB observation will provide a solid measurement of the overall merger rate densities of MBHBs, with other model parameters being constrained to different degrees depending on the array. We found a strong degeneracy between e_t and the typical mass scale of merging MBHBs, defined by the parameter $\mathscr{M}_*$. The degeneracy can be broken only with a confident detection of the high frequency drop of the spectrum, which depends on the mass underlying mass function but not on the eccentricity at decoupling. Unfortunately, this is possible only if the signal is detected at $f \gtrsim \text{few} \times 10^{-8}$Hz, which might be out of range even for the SKA. Finally, we considered the benefit of PTA detection when priors on the MBHB mass functions provided by independent observations are folded into the analysis. We found that, in this case, even in the *IPTA30* case, the eccentricity parameter can be constrained, because the constrained prior allows a better measurement of the typical MBHB mass scale. Therefore, when combined with independent observations, PTA observations in the foreseeable future have the potential of greatly enhancing our knowledge of MBHB astrophysics and dynamics.

These results are subject to a number of caveats, that will be explored in future work. First, we did not considered measurement errors in the observations. Although we included uncertainties in the measured characteristic amplitude at each frequency, we centred them at the value of the injected signal. Including an additional scatter, will make the reconstruction of the spectrum more cumbersome, especially in the case of low S/N detection. Second, we did not include the intrinsic scatter of the signal amplitude due to the stochastic nature of the GWB. In our model, each set of parameters produces a single $h_c(f)$. However, the exact value of the GWB at each frequency depends on the statistics of rare massive systems, and therefore, each set of underlying MBHB population parameters produced a probability distribution of $h_c(f)$ at each frequency. This can be taken into account with a suitable modification of the likelihood function that we plan to implement as next step of this investigation. Finally, our current analysis is limited to the stochastic part of the signal. Especially at high frequency, bright sources will be individually resolvable, carrying a great deal of information about the most massive systems that can be used to complement the information provided by the GWB spectral shape. All these shortcomings can be addressed within our framework via suitable modifications of different stages of the pipeline, and will be the subject of future publications in this series of papers.

ACKNOWLEDGEMENTS

We acknowledge the support of our colleagues in the European Pulsar Timing Array. A.S. is supported by a University Research Fellow of the Royal Society.

REFERENCES

Abbott B. P., et al., 2016a, Physical Review Letters, 116, 061102
Abbott B. P., et al., 2016b, Physical Review Letters, 116, 221101
Abbott B. P., et al., 2016c, Physical Review Letters, 116, 241103
Arzoumanian Z., et al., 2016, Astrophysical Journal, 821, 13
Babak S., et al., 2016, MNRAS, 455, 1665
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Booth R. S., de Blok W. J. G., Jonas J. L., Fanaroff B., 2009, preprint, (arXiv:0910.2935)
Chen S., Sesana A., Del Pozzo W., 2016, preprint, (arXiv:1612.00455)
Croton D. J., et al., 2006, MNRAS, 365, 11
Dabringhausen J., Hilker M., Kroupa P., 2008, MNRAS, 386, 864
Dehnen W., 1993, MNRAS, 265, 250
Del Pozzo W., Veitch J., 2015, CPNest: Parallel nested sampling in python, https://github.com/johnveitch/cpnest
Desvignes G., et al., 2016, MNRAS, 458, 3341
Dotti M., Sesana A., Decarli R., 2012, Advances in Astronomy, 2012, 3
Ellis J. A., Siemens X., Creighton J. D. E., 2012, Astrophysical Journal, 756, 175
Enoki M., Nagashima M., 2007, Progress of Theoretical Physics, 117, 241
Foster R. S., Backer D. C., 1990, Astrophysical Journal, 361, 300
Hellings R. W., Downs G. S., 1983, The Astrophys. J.l, 265, L39
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, ApJS, 163, 1

Hopkins P. F., Richards G. T., Hernquist L., 2007, Astrophysical Journal, 654, 731
Kelley L. Z., Blecha L., Hernquist L., 2016, MNRAS,
Kocsis B., Sesana A., 2011, MNRAS, 411, 1467
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Lentati L., Alexander P., Hobson M. P., Taylor S., Gair J., Balan S. T., van Haasteren R., 2013, Phys. Rev. D, 87, 104021
Lentati L., et al., 2015, MNRAS, 453, 2576
Lin L., et al., 2004, Astrophysical Journal, 617, L9
Lotz J. M., Jonsson P., Cox T. J., Croton D., Primack J. R., Somerville R. S., Stewart K., 2011, Astrophysical Journal, 742, 103
McWilliams S. T., Ostriker J. P., Pretorius F., 2014, Astrophysical Journal, 789, 156
Middleton H., Del Pozzo W., Farr W. M., Sesana A., Vecchio A., 2016, MNRAS, 455, L72
Moore C. J., Taylor S. R., Gair J. R., 2015, Classical and Quantum Gravity, 32, 055004
Nan R., et al., 2011, International Journal of Modern Physics D, 20, 989
Petiteau A., Babak S., Sesana A., de Araújo M., 2013, Phys. Rev. D, 87, 064036
Phinney E. S., 2001, ArXiv Astrophysics e-prints,
Rasskazov A., Merritt D., 2016, preprint, (arXiv:1606.07484)
Ravi V., Wyithe J. S. B., Shannon R. M., Hobbs G., Manchester R. N., 2014, MNRAS, 442, 56
Reardon D. J., et al., 2016, MNRAS, 455, 1751
Rosado P. A., Sesana A., Gair J., 2015, MNRAS, 451, 2417
Sesana A., 2013a, Classical and Quantum Gravity, 30, 224014
Sesana A., 2013b, MNRAS, 433, L1
Sesana A., Vecchio A., Colacino C. N., 2008, MNRAS, 390, 192
Sesana A., Vecchio A., Volonteri M., 2009, MNRAS, 394, 2255
Shankar F., Salucci P., Granato G. L., De Zotti G., Danese L., 2004, MNRAS, 354, 1020
Shannon R. M., et al., 2015, Science, 349, 1522
Simon J., Burke-Spolaor S., 2016, Astrophysical Journal, 826, 11
Skilling J., 2004, in American Institute of Physics Conference Series. pp 395–405
Springel V., et al., 2005, Nature, 435, 629
The NANOGrav Collaboration et al., 2015, Astrophysical Journal, 813, 65
Veitch J., Vecchio A., 2010, Phys. Rev. D, 81, 062003
Veitch J., et al., 2015, Phys. Rev. D, 91, 042003
Verbiest J. P. W., et al., 2016, MNRAS, 458, 1267
Volonteri M., Haardt F., Madau P., 2003, Astrophysical Journal, 582, 559
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Xu C. K., Zhao Y., Scoville N., Capak P., Drory N., Gao Y., 2012, Astrophysical Journal, 747, 85
de Ravel L., et al., 2009, A&A, 498, 379

MNRAS **000**, 1–1 (2017) Preprint 22 September 2017 Compiled using MNRAS LATEX style file v3.0

Erratum: Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays

Siyuan Chen,[1]⋆ Hannah Middleton,[1]† Alberto Sesana,[1] Walter Del Pozzo[1,2] and Alberto Vecchio[1]
[1]*School of Physics & Astronomy, University of Birmingham, Birmingham, B15 2TT, UK*
[2]*Dipartimento di Fisica "Enrico Fermi", Università di Pisa, Pisa I-56127, Italy*

Accepted . . . Received . . . ; in original form . . .

It has been brought to our attention that there is an inconsistency between the content of figure 5 of the original paper (Chen et al. 2017) and the related statement that the figure shows a simulation of a gravitational wave background generated by binaries with eccentricity of 0.01. After a careful investigation we realized that the original figure 5 was indeed produced using a simulation with binaries with an eccentricity of 0.1, consistent with the recovered value. We would like to apologize for any confusion that figure has caused. We show below the correct figure produced with an $e_t = 0.01$ injection. The analysis correctly recovers the injected value and the peak around $e_t = 0.1$ (present in the original figure) disappears, as expected.

ACKNOWLEDGEMENTS

We acknowledge the support of our colleagues in the European Pulsar Timing Array. A.S. is supported by a University Research Fellow of the Royal Society.

REFERENCES

Chen S., Middleton H., Sesana A., Del Pozzo W., Vecchio A., 2017, MNRAS, 468, 404

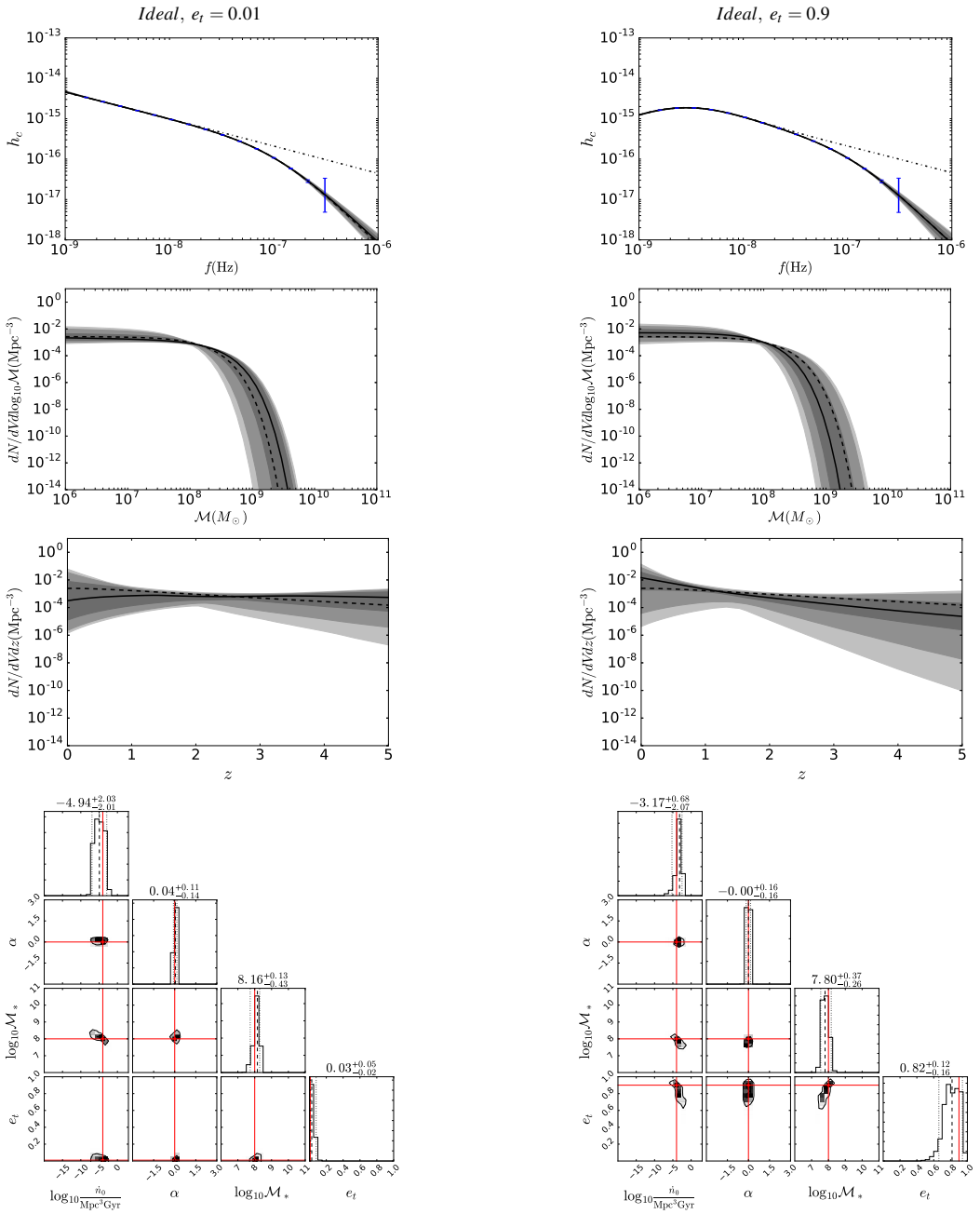


Figure 1. Implication of an ideal detection with 500 MSPs timed at sub-ns precision for 30 years. The injected model has default parameters with $e_t = 0.01$ (left column) and $e_t = 0.9$ (right column). Panel sequence and style as in figure 4 [of the original paper].