Evincing the histories of the cosmic supermassive black hole and galaxy populations with gravitational waves

By

Vikram Ravi

Submitted in total fulfilment of the requirements
of the degree of Doctor of Philosophy

February 2015

School of Physics
The University of Melbourne

Produced on archival quality paper.

Evincing the histories of the cosmic supermassive black hole and galaxy populations with gravitational waves

Abstract

Evincing the histories of the cosmic supermassive black hole and galaxy populations with gravitational waves

By Vikram Ravi

Supermassive black holes (SMBHs) are inferred to exist at the centres of massive galaxies throughout the Universe. When two such galaxies merge, a binary SMBH system is likely to form, which coalesces following losses of energy and angular momentum to gravitational waves (GWs). GWs from binary SMBHs will cause metric perturbations at the Earth that affect the arrival times of pulses from radio pulsars within our Galaxy. This concept has led to the establishment of pulsar timing array (PTA) collaborations, which are primarily aimed at detecting GWs. This thesis is motivated by the prospect of gleaning insights into the assembly histories of the cosmological SMBH and galaxy populations by searching for GWs from binary SMBHs.

I show that a GW background (GWB) generated by the binary SMBH population is the most promising class of signal to consider for this purpose, with the strong possibility of a detection within the forthcoming decade. In contrast, I find that GWs from individual binary SMBHs are not viable sources for current PTA searches. I also demonstrate that the statistics of pulsar timing variations induced by the GWB will be mildly non-Gaussian. By developing techniques to simulate the effects of GWs from predicted populations of binary SMBHs on PTA data, I find that these non-Gaussian statistics result in a $\sim 10\%$ degradation in the recent Parkes PTA upper limit on the GWB. In a separate investigation, I show that interactions between binary SMBHs and their environments may cause attenuation in the GWB at frequencies up to $10^{-8}\,\mathrm{Hz}$. Finally, upon comparing various predictions for the GWB with the most recent upper limits from the Parkes PTA, I find that a model which posits purely merger-driven growth of massive galaxies during the last 8 billion years is excluded at the 91% confidence level. I also derive a constraint on the merger timescale of massive galaxies: I find that the mean time spent between projected separations of 20 and $5\,h^{-1}\,\mathrm{kpc}$ is greater than $0.1\,\mathrm{Gyr}$ with 95% confidence.

This is to certify that:

1. the thesis comprises only my original work towards the PhD except where indicated in the Preface,
2. due acknowledgement has been made in the text to all other material used,
3. the thesis is fewer than 100 000 words in length, exclusive of tables, maps, bibliographies and appendices.

Candidate Date

Preface

The doctoral research that I present in this thesis was conducted between March 2011 and September 2014, under the joint supervision of Prof. Stuart Wyithe at the University of Melbourne and Dr. George Hobbs at CSIRO Astronomy and Space Science. During this time, I have been a member of the Parkes Pulsar Timing Array (PPTA) collaboration, first led by Prof. Richard Manchester and more recently by Dr. Hobbs. The PPTA comprises scientists and technical staff from Australia, the United States of America, China and Germany. My research contributed directly to a central goal of the PPTA, which is to perform astrophysics through searches for gravitational waves in radio pulsar timing data.

Throughout this thesis, I endeavour to distinguish between studies that I carried out and results that, while I may have contributed to, I am not directly responsible for. I led a selection of published works during my doctorate, and was a co-author of others. The publications of direct relevance to this thesis include:

1. Ravi, V., et al. 2012, *Does a "Stochastic" Background of Gravitational Waves Exist in the Pulsar Timing Band?*, ApJ, 761, 84

2. Manchester, R. N., (21 authors), Ravi, V., (8 authors) 2013, *The Parkes Pulsar Timing Array Project*, PASA, 30, 17

3. Shannon, R. M., Ravi, V. (co-first author), et al. 2013, *Gravitational-wave limits from pulsar timing constrain supermassive black hole evolution*, Science, 342, 334

4. Ravi, V. et al. 2014, *Binary supermassive black hole environments diminish the gravitational wave signal in the pulsar timing band*, MNRAS, 442, 56

5. Zhu, X.-J., (15 authors), Ravi, V., (2 authors) 2014, *An all-sky search for continuous gravitational waves in the Parkes Pulsar Timing Array data set*,

MNRAS, 444, 3709

6. Wang, J.-B., (6 authors), Ravi, V., (12 authors) 2015, *Searching for gravitational wave memory bursts with the Parkes Pulsar Timing Array*, MNRAS, 446, 1657

7. Ravi, V., et al. 2015, *Prospects for gravitational-wave detection and supermassive black hole astrophysics with pulsar timing arrays*, MNRAS, 447, 2772.

I am responsible for the majority of the contents of publications (1), (4) and (7) listed above; Chapters 5, 6 and 7 of this thesis correspond primarily to these respective publications. Section 3.1.4 in Chapter 3 also contains material from publication (4). In addition, I was a co-first author for publication (3) listed above. Parts of Chapters 5 and 8 are based on my contributions to publication (3). In Chapter 4, I summarise the results of publications (2), (3), (5) and (6). The remainder of the material presented in this thesis, that is, Chapter 1, Chapter 2 and most of Chapter 3, forms a review of background concepts and does not represent original research.

I begin each of Chapters 2−8 with summary paragraphs displayed in bold face. Shorter summaries of each Chapter may be found at the end of Chapter 1.

Acknowledgments

My first thanks are to my PhD supervisors: Prof. Stuart Wyithe and Dr. George Hobbs. Stu very kindly took me on as a student based on an e-mail proposing a project out of the blue, and has been a source of unbelievably rapid insights, masterful academic help, and inspiration. While my first meeting with George may have involved me teaching him how to play table tennis, the tables have since turned in other respects. In addition to being a great friend and enthusiastic mentor, George has taught me most of what I know about the practice of observational science and academic communication, and never ceases to amaze with his lightning problem-solving abilities.

Dr. Ryan Shannon has effectively been a third supervisor, and his contributions to guiding my research cannot be overstated. I hope that our collaboration on many, many things, and our luck (?!), can continue well into the future. I thank the various people involved with the Parkes Pulsar Timing Array collaboration, in particular Profs. Dick Manchester and Yuri Levin and Dr. Sarah Burke-Spolaor for their strong interest in and support of my work. The opportunity to discuss my research with members of the wider International Pulsar Timing Array collaboration has also been invaluable. Drs. Alberto Sesana and Rutger van Haasteren have provided extremely useful (and sometimes sudden!) feedback.

The University of Melbourne Astronomy Group has been my academic home for the last three and a half years. I thank Dr. Paul Lasky for numerous discussions on topics related and unrelated to my thesis, and Prof. Andrew Melatos for sharing some of his gargantuan physical insight and forming part of my doctoral advisory committee. I thank Dr. Simon Mutch for lending his hard-earned experience with semi-analytic galaxy formation models to this thesis. To the Melbourne Uni astro students, you were examples of academic rigour (seriously!), and taught me a great deal about practical applications of statistics. The breadth of my astro-

physical knowledge was enjoyably increased through various journal clubs, including cosmojam and the neutron stars / compact objects / pulsars / gravitational waves meetings.

I am particularly grateful to CSIRO Astronomy and Space Science for providing me with a studentship, and for supporting my innumerable (and at times hastily organised) visits to Marsfield, Parkes and Narrabri. I would like to extend my thanks to the CSIRO Astronomy and Space Science operations staff at all sites. I acknowledge the support of the Science and Industry Endowment Fund in providing me with a John Stocker Postgraduate Scholarship. I was also the recipient of an Australian Postgraduate Award, funded by the Australian Government. This research makes use of data from the Parkes telescope, which is part of the Australia Telescope funded by the Commonwealth Government for operation as a National Facility managed by CSIRO.

The Millennium and Millennium-II simulation data bases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory. Much of my work was performed on the swinSTAR supercomputer at the Swinburne University of Technology. This dissertation was typeset using the ucastrothesis LaTeX template.

Special thanks go to Du Sha and her family for taking care of me in Beijing, and for accommodating my demands on George's time. Jonathan Khoo gets there with a shout-out for heaps of stuff. To Emma, my partner and closest friend throughout my PhD, and her family, thank you so much for the love, support and patience you have shown a crazy astronomer. My final thanks go to Mama and Papa: thank you for making me who I am today.

Contents

List of commonly-used abbreviations

A&A Astronomy and Astrophysics.

AGN Active Galactic Nucleus.

AJ The Astronomical Journal.

ApJ The Astrophysical Journal.

ApJL The Astrophysical Journal Letters.

ApJS The Astrophysical Journal Supplement.

ARA&A Annual Reviews in Astronomy and Astrophysics.

CMB Cosmic Microwave Background.

EPS Extended Press-Schechter.

EPTA European Pulsar Timing Array.

FAST Five hundred metre Aperture Spherical Telescope.

GSMF Galaxy Stellar Mass Function.

GW Gravitational Wave.

GWB Gravitational-Wave Background.

IPTA International Pulsar Timing Array.

ΛCDM Lambda − Cold Dark Matter.

MNRAS Monthly Notices of the Royal Astronomical Society.

NANOGrav North American Nanohertz Observatory for Gravitational waves.

PASA Publications of the Astronomical Society of Australia.

PASP Publications of the Astronomical Society of the Pacific.

PPTA Parkes Pulsar Timing Array.

PTA Pulsar Timing Array.

SFR Star Formation Rate.

SKA Square Kilometre Array.

SMBH Supermassive Black Hole.

ToA Time of Arrival.

Chapter 1

Introduction

Through careful observation of the night sky, the earliest astronomers the world over realised that the Earth is most closely surrounded by a Solar System, comprised of planets and our Sun. Beyond the Solar System lay the fixed stars, apparently so for eternity. However, despite the dogmatic promulgation of such Aristotelian ideas, it was clear to most ancient civilisations that the stars were *not* unchangeable. As early as 1000 BC, Egyptian scholars producing calendars of prognoses for lucky and unlucky days may have accurately measured the $\sim 2.85\,\mathrm{day}$ period of the eclipsing binary system Algol (Jetsu et al. 2013). Ancient Chinese astronomers recorded numerous supernova explosions, the first known of which occurred in 185 AD, which they termed 'guest stars' and identified as distinct from the planets. With the discovery of stellar proper motions by Halley, who compared his astrometric measurements to those of Hipparchus obtained more than 1800 years previously, even the supposed stationary nature of the stars was disproved. Today, our vision of the Universe is transformed from the somewhat forbidding view of an eternal, static firmament, to one of unimaginable dynamism. The Earth orbits the Sun with a speed of $30\,\mathrm{km\,s^{-1}}$, the Sun orbits the Milky Way at $220\,\mathrm{km\,s^{-1}}$, and the Milky Way is approaching the nearby Andromeda spiral galaxy at $110\,\mathrm{km\,s^{-1}}$. Neutron stars, consisting of material weighing more than our Sun compressed into objects a few tens of kilometres in diameter, rotate up to once every millisecond; radiation and particle jets powered by massive black holes at the centres of galaxies may vary in intensity in just a few hours; gamma-ray bursts at distances of 300 billion light years outshine entire galaxies for just a few seconds. The expansion of space since the beginning of the Universe causes all distant galaxies to appear to be receding

from us at ever-increasing speeds.

Interwoven with this thesis is a sense of the Universe changing appreciably on timescales that we can comprehend. This thesis concerns the expected gravitational-wave signals from binary black holes at the centres of merging galaxies throughout the Universe and the search for these signals with observations of pulsars within the Milky Way. Almost all galaxies are thought to host central black holes. When a pair of galaxies merges, the black holes will form a binary system. Binary black holes lose energy and angular momentum to gravitational waves, which are periodic perturbations to the geometry of space that travel at the speed of light. Gravitational waves from binary black holes with few-year orbital periods in distant galaxies, as they propagate through the Milky Way, cause apparent errors in timing measurements of radio pulsars. Pulsars, which are neutron stars with fixed beams of radio waves that sweep by the Earth during every millisecond- to second-long rotation, are extraordinary celestial clocks; models for the pulse arrival times measured using radio telescopes are precise to better than 100 ns over a decade. By searching for time-variable timing errors that are correlated between different pulsars, the passage of gravitational waves past the Earth may be discovered. This concept may be implemented as a pulsar timing array.

My emphasis on gravitational waves as a means of investigating the Universe moves this thesis outside the traditional, evolutionary stream of astrophysics. Given that astro-particle science with cosmic rays and neutrinos is still in its infancy, essentially all astrophysical knowledge is founded on observations of electromagnetic radiation over 19 orders of magnitude in wavelength. Gravitational waves are most strongly generated by the motions of extreme concentrations of mass, such as supernovae, or of gravitationally-bound systems of compact objects such as white dwarfs, neutron stars or black holes. The compactness of such objects, or the lack of detectable electromagnetic radiation in the case of black holes, can hinder their characterisation through electromagnetic observations. Despite evidence that the orbits of binary pulsar systems decay at rates corresponding to the emission of gravitational waves (e.g., Hulse & Taylor 1975), signatures of *propagated* gravitational radiation have not yet been detected. However, large-scale experiments over the next decade, such as pulsar timing arrays, should see this barrier overcome, shedding a new light on some of the most fascinating astrophysical events and systems.

In March 2011, when I began the investigations presented in this thesis, the

first data releases from modern, dedicated pulsar timing array implementations were being assembled. Three pulsar timing array collaborations[1] had been conducting observations of millisecond pulsars since 2004, and had just formed the International Pulsar Timing Array consortium (Hobbs et al. 2010b). The primary gravitational-wave signal of interest to pulsar timing array collaborations was a stochastic, isotropic background generated by the predicted cosmological population of binary black holes (Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003a; Enoki et al. 2004; Sesana et al. 2008b). Jenet et al. (2005) had shown that weekly observations of 20 pulsars for 10 yr that attained 100 ns timing precisions for each pulsar could lead to the detection of such a gravitational-wave background, with a signal to-noise-ratio of ≥ 4. Further theoretical investigations by Sesana et al. (2009) suggested that gravitational waves from individual binary black holes may also be detectable with pulsar timing arrays.

The central rationale for this thesis is the prospect of gravitational-wave constraints or detections from pulsar timing arrays furthering knowledge of how the cosmological populations of black holes and galaxies formed and evolved. At the beginning, I was motivated in particular by the upcoming releases and analyses of pulsar timing data sets that promised to be vastly more sensitive to gravitational waves than ever before. While pulsar timing arrays have not yet achieved a detection of gravitational waves, this promise has nonetheless come to fruition. For example, current upper limits on the energy density of the gravitational-wave background from binary black holes (Shannon et al. 2013) are six times lower than previous constraints (van Haasteren et al. 2011).

Important science is possible prior to a "direct" detection of gravitational radiation. In particular, astrophysical models may be identified which predict gravitational-wave signals that would most likely be perceivable in data from current detection experiments. In this thesis, I implement this concept by systematically characterising expected gravitational-wave signals given plausible scenarios for the cosmological population of binary black holes formed in galaxy mergers. I then compare these predictions with the latest pulsar timing array data. Besides the innate scientific importance of this study, the theoretical predictions I present will serve to

[1]These included the European Pulsar Timing Array (EPTA; Kramer & Champion 2013), the North American Nanohertz Observatory for Gravitational waves (NANOGrav; McLaughlin 2013) and the Parkes Pulsar Timing Array (PPTA; Manchester et al. 2013); all continue to be active and have grown substantially since 2011. I am a member of the PPTA collaboration.

guide searches for gravitational waves with pulsar timing arrays.

1.1 Outline of this thesis

Chapters 2 and 3 form a review of concepts underlying this thesis, and Chapter 4 summarises the best existing gravitational-wave constraints from pulsar timing arrays. Chapters 5, 6, 7 and 8 describe my original research. I present my conclusions in Chapter 9. More detailed outlines of each Chapter are as follows.

Chapter 2. I review current knowledge of the cosmology and astrophysics underlying the formation and growth of galaxies and black holes. I focus in particular on concepts that directly determine predictions for gravitational waves from binary black holes, and discuss some unsolved problems of relevance to this thesis.

Chapter 3. I review the body of primarily theoretical studies on the formation of binary black holes and their gravitational-wave emission, and summarise the technical aspects of pulsar timing arrays. Some original work on defining the gravitational-wave background and its relation to a cosmological source population may be found in §3.1.4.

Chapter 4. I review the best existing pulsar timing array constraints on gravitational-wave signals from binary black holes. I describe an upper limit on the strength of the stochastic gravitational-wave background (Shannon et al. 2013), along with upper limits on the all-sky occurrence of individual binary black hole gravitational-wave sources (Zhu et al. 2014; Wang et al. 2015).

Chapter 5. I present my investigation into the statistics of the pulsar timing errors induced by gravitational waves from the cosmological binary black hole population. Previous studies had assumed that these statistics corresponded to a Gaussian random process. By employing a physically-motivated model for binary black holes, I show that this is not entirely correct, because of the likely presence of a few, particularly massive and nearby binary systems. I demonstrate that this marginally reduces our ability to constrain the overall strength of the gravitational-wave background from binary black holes.

Chapter 6. I present my assessment of the likely effects of binary black hole environments and eccentricities on the gravitational-wave background, as well as the potential for detectable bursts of gravitational waves from close approaches between black holes. I show that the lowest-frequency gravitational waves may be attenuated in comparison with previous studies.

Chapter 7. I present my investigation into what ranges of gravitational-wave signals are expected from binary black holes given empirical constraints from electromagnetic observations. I find that while a stochastic gravitational-wave background is likely to be detectable with pulsar timing arrays within the next decade, gravitational waves from individual binary black holes will only be detectable with next-generation experiments.

Chapter 8. I compare a range of scenarios for gravitational waves from binary black holes with the best existing pulsar timing array results summarised in Chapter 4. I find that certain scenarios are already inconsistent with current measurements.

Chapter 9. I present my main conclusions from this thesis, as well as my prognosis for the future of this research.

Chapter 2

Galaxy and supermassive black hole formation and growth

I review the cosmology and astrophysics behind the formation and growth of galaxies and supermassive black holes (SMBHs). These concepts directly determine predictions for gravitational waves (GWs) from binary SMBHs, and this domain of astrophysical knowledge may also be furthered by GW constraints. I consider in particular the hierarchical growth of dark matter structures, the formation of the first galaxies and SMBHs, and the co-evolution of the present-day populations of galaxies and SMBHs. I also discuss some unsolved problems of direct relevance to this thesis.

The gravitationally bound systems of stars and gas called galaxies are the fundamental observed building blocks of our Universe. Galaxies of myriad morphologies and sizes exist in the immediate surrounds of our own Milky Way (e.g., Hubble 1926a), and galaxies with similar properties have been observed throughout cosmic time (for a review, see Conselice 2014). In the nearby Universe, galaxies with stellar masses greater than $10^{11} M_{\odot}$ are typically 'elliptical', or spheroidal (somewhat contradictorily), in morphology, and galaxies with stellar masses in the range $10^{9} M_{\odot} - 10^{11} M_{\odot}$ are typically disk-shaped with 'spiral' morphologies (Cappellari et al. 2011). Elliptical galaxies have old, redder stellar populations, without active star formation and with no gas reservoirs for future star formation. There

is little order to the stellar orbits. Disk-shaped galaxies, on the other hand, may either consist primarily of dust and old stars (such systems are termed 'lenticulars', or S0s), or younger and bluer stars and gas ordered into dense spiral arms with clumps of star formation activity. Ellipticals and S0s are referred to as 'early-type' galaxies, and spiral systems are termed 'late-type' galaxies after their locations within the Hubble (1936) 'tuning fork' galaxy morphological classification scheme. More massive, spiral galaxies like the Andromeda galaxy M31 (with a stellar mass of $(1.04 \pm 0.05) \times 10^{11} M_{\odot}$, Geehan et al. 2006) typically have central 'bulges', or spheroidal distributions of older stars. These spheroids are, in essence, scaled-down versions of elliptical galaxies. Dwarf galaxies with masses below $10^{9} M_{\odot}$ have irregular or spheroidal structures, with a range of stellar populations. A summary of various salient properties of elliptical, lenticular, spiral and irregular galaxies is shown schematically in Figure 2.1.

Massive galaxies in the local Universe appear to ubiquitously host central supermassive black holes (SMBHs; Kormendy & Ho 2013), with masses in the range $10^{6} M_{\odot} - 10^{11} M_{\odot}$. Black holes are collapsed objects with their mass enclosed by an event horizon, which, in the non-rotating case, has a radius (Schwarzschild 1916):

$$R_S = \frac{2GM_{\bullet}}{c^2} = 2.95 \frac{M_{\bullet}}{M_{\odot}} \,\mathrm{km} \tag{2.1}$$

where R_S denotes the Schwarzschild radius, G is the universal gravitational constant, $M_{\bullet}$ is the mass of the black hole and c is the vacuum speed of light. An event horizon can only be crossed in the inwards direction, although the dilation of time in strong gravitational fields and the resultant redshifting of emitted light implies that nothing can ever be seen to actually cross the horizon. Evidence for astrophysical black holes was first identified through observations of active galactic nuclei (AGN), or quasars (Salpeter 1964; Zel'dovich 1964), which are now interpreted as SMBHs that are surrounded by hot infalling gas and which may power fantastically energetic jets and outflows. Specific evidence for SMBHs as quasar engines comes from their large luminosities given their cosmological distances (e.g., Schmidt 1963), and small sizes implied by rapid variability across the electromagnetic spectrum (e.g., Matthews & Sandage 1963; Ulrich et al. 1997). The space density of quasars at higher redshifts and their inferred short duty cycles led Lynden-Bell (1969) and Lynden-Bell & Rees (1971) to predict the existence of SMBHs at the centres of

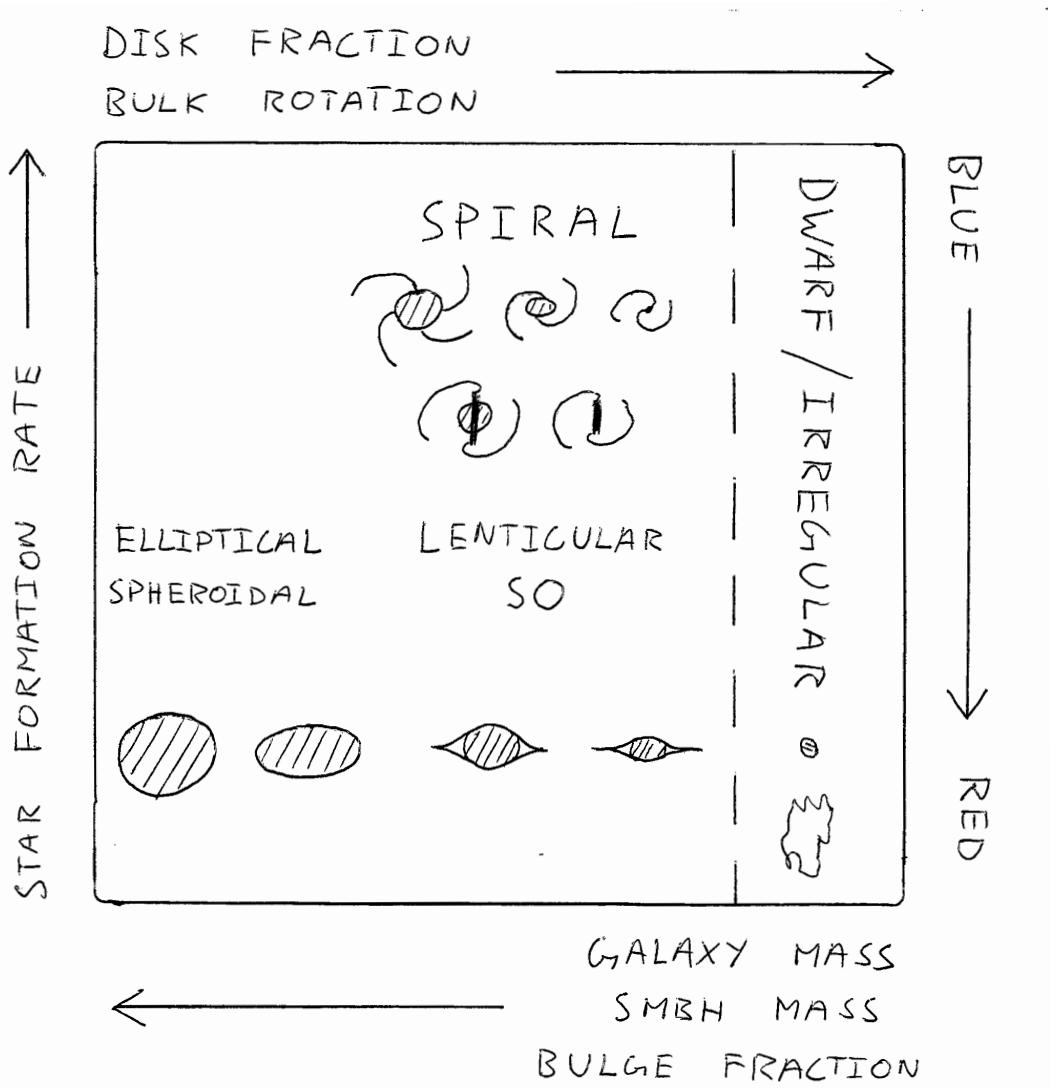

Figure 2.1: A galaxy classification scheme based on the 'comb' of Cappellari et al. (2011), illustrating the relative properties of elliptical, lenticular (S0), spiral and dwarf/irregular galaxies. Schematic illustrations of elliptical and lenticular galaxies are drawn at the bottom of the figure, schematic illustrations of spiral galaxies are drawn at the top of the figure, and schematic illustrations of dwarf/irregular galaxies are drawn at the bottom right of the figure. Please note that the location indicated on the figure for dwarf/irregular galaxies is not wholly accurate: such systems also exhibit a large range of bulge fractions, disk fractions and bulk rotations.

galaxies in the local Universe, including the Milky Way.

The presence of a central quiescent SMBH in a galaxy is most commonly deduced through dynamical modelling of observations of the kinematics of stars and gas with sufficient spatial resolution. SMBHs have been discovered and their masses measured using these techniques at the centres of 87 nearby galaxies at the time of writing (Kormendy & Ho 2013). In our Milky Way, observations of individual stellar orbits (e.g., Ghez et al. 2008) constrain the mass of the SMBH associated with the bright radio source Sagittarius A* (Balick & Brown 1974) to be $(4.1 \pm 0.6) \times 10^6 M_\odot$. In other galaxies, typical $\sim 68\%$ confidence intervals of ~ 0.3 dex are attained for SMBH mass measurements through observations of stellar or ionised gas kinematics and brightness distributions, while observations of water maser disks yield confidence intervals that are < 0.05 dex wide (Kuo et al. 2011). The heaviest SMBHs are found in brightest cluster galaxies (BCGs); the most massive SMBH known, in the BCG (NGC 4889) of the Coma galaxy cluster, has a mass in the range $9.8 \times 10^9 M_\odot - 2.1 \times 10^{10} M_\odot$ (McConnell et al. 2011). Mass measurements of nearby SMBHs that are AGN are also possible using reverberation mapping (for details, see Peterson 1993).

The masses of SMBHs hosted within galaxy bulges, which are either the entirety of elliptical galaxies or parts of S0 and spiral galaxies, exhibit tight log-linear correlations with various bulge properties (for a review, see Kormendy & Ho 2013). These include the bulge velocity dispersions (the well-known $M_\bullet - \sigma_c$ relation), luminosities and masses; the approximately linear relation between SMBH masses and bulge masses ($M_{\rm bul}$) results in a typical ratio $M_\bullet / M_{\rm bul} \sim 0.005$. Together with further striking correlations between SMBH masses and the apparent central stellar density deficits or 'cores' in most massive elliptical galaxies (e.g., Graham 2012), these relations point to a fundamental link between the growth processes of spheroids and SMBHs. Such a link is remarkable given the order-of-magnitude differences in scale between the gravitational influence regions of SMBHs and the sizes of their host bulges.

The galaxy population was significantly different at earlier cosmological epochs as compared to the present day. At earlier times, fewer massive elliptical galaxies were present, and a larger fraction of massive galaxies had irregular morphologies and high star formation rates (SFRs). The cosmic SFR density peaked at a redshift of $z \sim 2$, corresponding to a look-back time of $\sim 10^{10}$ yr; this rate was an order of magnitude

higher than the current value (Hopkins & Beacom 2006). At these redshifts, the massive galaxy population was evenly split between elliptical, spiral and irregular galaxies, and approximately 20% of galaxies were undergoing mergers with other galaxies at a given time (Conselice 2014). Our most distant galaxy observations reveal an abundance of blue massive systems with extreme SFRs. Indeed, a recent observation of a galaxy at $z = 7.5$ with a mass of $\sim 10^9 M_\odot$ revealed a SFR more than a hundred times greater than that of the Milky Way (Finkelstein et al. 2013).

The AGN population also evolves with cosmic time (e.g., Alexander & Hickox 2012). The evolution of the space density of luminous quasars with redshift approximately traces the evolution of the SFR density, with a peak at $z \sim 2$. The typical quasar bolometric luminosity also appears to increase with redshift (Hopkins et al. 2007). Various authors have employed the Soltan argument (Soltan 1982) to show that the local space density of SMBHs is consistent with the mass accreted during short-lived ($\sim 10^8$ yr) luminous quasar phases as traced by the observed bolometric quasar luminosity function (e.g., Chokshi & Turner 1992; Yu & Tremaine 2002; Shankar et al. 2013). It is particularly remarkable that luminous quasars, potentially hosting SMBHs with masses greater than $10^9 M_\odot$, have been observed at redshifts as high as 7.1 (Fan 2006; Mortlock et al. 2011), less than 10^9 yr after the Big Bang.

Identifying and synthesising the physical processes underpinning the formation and growth of galaxies is a core task of astrophysics. To this end, understanding the evolution of the SMBH population throughout the history of the Universe appears to be crucial. The study of phenomena associated with SMBHs also allows for fundamental physical insights in regimes not accessible to Earth-bound laboratories. In the following sections, I will summarise our physical understanding of the formation and co-evolution of galaxies and SMBHs. I will begin, in §2.1, by describing the evolving dark matter distribution that dominates the gravitational dynamics of all galaxies. I will then consider the origins of the first galaxies and SMBHs in §2.2, and progress to the present-day by outlining current knowledge of the assembly of the galaxy and SMBH populations in §2.3. I will conclude this Chapter in §2.4 with a summary of some unsolved problems that motivate this thesis.

2.1 Structure formation: the assembly of dark matter haloes

2.1.1 The existence of dark matter, and cosmic inflation

Since the realisation (Opik 1922; Hubble 1925, 1926b) that our Universe consists of a homogeneous and isotropic galaxy distribution, various attempts have been made to measure the mean cosmological mass density. Using only a selection of luminous nearby galaxies, Hubble (1926b) found a value of $1.5 \times 10^{-28}\,\mathrm{kg\,m^{-3}}$, roughly a factor of three below current estimates for the baryon density. It was quickly realised, however, that relating the luminosities and velocity dispersions of galaxies to their masses, and indeed to the cosmological mass density, was fraught with difficulty (Zwicky 1937). Instead, Zwicky (1933, 1937) and Smith (1936) argued for the use of the virial theorem as applied to galaxy clusters to place lower limits on the cluster masses using the measured line-of-sight velocity dispersions as follows:

$$M_{\mathrm{cluster}} > \frac{3 R_{\mathrm{cluster}} \sigma^2_{\mathrm{LOS}}}{5G}, \tag{2.2}$$

where R_{cluster} is the cluster radius and σ_{LOS} is the line-of-sight velocity dispersion (which, strictly speaking, should be mass-weighted). When Zwicky (1933) estimated the mass of the Coma cluster using this technique, the result was 400 times larger than the estimated density of the luminous cluster matter.[1] Zwicky (1933) concluded, both for the Coma cluster and the Universe as a whole (my translation): "if this surprising result should be true, dark matter is present in much greater density than luminous matter."

Numerous lines of evidence (see, e.g., van den Bergh 1999; Einasto 2013) now exist for the paradigm that only $\sim 16\%$ of the matter density of the Universe exists as "baryons" of the forms described by the Standard Model of particle physics (Planck Collaboration et al. 2013). The early findings on cluster masses (Zwicky 1933; Smith 1936; Zwicky 1937) were supplemented with the realisation (e.g., van den Bergh 1962) that without the high inferred mass-to-light ratios clusters could not be dynamically stable over timescales inferred from the ages of the member galaxy stellar populations. Similar fractions of "missing mass" were inferred

[1] The factor of 400 should have been a factor of ~ 50 because of the assumption of a Hubble constant of $558\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

from the dynamics of samples of galaxy pairs (e.g., Page 1960), with the alternative hypothesis again being ubiquitous positive total energies among these systems. Within individual galaxies, measurements of the bulk rotations of galactic disks revealed constant rotation velocities as functions of disk radii, which were inconsistent (assuming Newtonian dynamics) with the inferred mass-density profiles of the luminous matter (e.g., Babcock 1939; Rubin & Ford 1970; Roberts & Whitehurst 1975; Roberts 1976). Various authors further showed that disks with the observed rotation curves and luminous matter density profiles would be dynamically bar-unstable (Ostriker & Peebles 1973), and that extended dark halos surrounding galaxies were a natural explanation for the observations (e.g., Ostriker et al. 1974). More recently, using a method first suggested by Zwicky (1937), modelling of the gravitationally-lensed light from distant galaxies by intervening galaxy clusters has further demonstrated the existence of dark matter (Kovner 1989; Massey et al. 2010; Hoekstra et al. 2013). Suggestions that a significant fraction of the inferred dark matter density in galactic haloes could lie in planetary-mass massive compact halo objects (MACHOs; Paczynski 1986) were shown to be unlikely by searches for the gravitational microlensing of Large Magellanic Cloud stars by intervening MACHOs (Alcock et al. 1996). More generally, given the realisation that baryonic dark matter candidates are unlikely to take any form besides MACHOs (e.g., Silk 1991), only the possibility of introducing non-baryonic dark matter candidates remained.

The origins of the idea that the formation of structure in the Universe was driven by the dynamics of dark matter are rooted in the motivations behind the theory of inflation (for a review, see Narlikar & Padmanabhan 1991). Chief among these motivations were the Dicke coincidences (Dicke 1970), which refer to:

The horizon problem. Regions of the Universe which were not causally connected in the past appear homogeneous; and

The flatness problem. A fine-tuned curvature parameter and matter density parameter are required at early times if curvature is present. Consider the Friedmann equation for the Hubble parameter $H = \dot{a}/a$, where a is the scale factor:

$$H^2 = \frac{8\pi G\rho_M}{3} + \frac{c\kappa}{a^2} + \frac{\Lambda}{3}, \tag{2.3}$$

where ρ_M is the matter density of the Universe, κ is the curvature parameter

and Λ is the cosmological constant. One can write $\Omega_M H^2 = \frac{8\pi G\rho}{3}$, where $\Omega_M = \rho_M/\rho_c$ is the ratio of the matter density to the critical density of the Universe. In the early 1980s, when inflation was introduced, it was generally believed that $\Lambda = 0$. Then, if $\Omega_M \neq 1$ and given then-existing constraints of $\Omega_M > 0.05$ (e.g., Peebles 1986), both the matter density and curvature contribute to the present-day value of H. However, given the different rates of evolution of the curvature and matter densities of the Universe with a (as $\sim a^{-2}$ and $\sim a^{-3}$ respectively), the curvature and matter densities would have to be wildly different at early times, but with specific values corresponding to our Universe. This would represent an undesirable fine-tuning scenario, which is solved by setting $\kappa = 0$.

To solve these and other problems such as the lack of evidence for magnetic monopoles (c.f. Narlikar & Padmanabhan 1991), Guth (1981) and Linde (1982) postulated that the Universe underwent a phase of exponential inflation at early times. A menagerie of physical explanations for inflation exist (Lyth & Riotto 1999), which generally invoke a phase transition of a universal quantum 'inflaton' field with negative pressure.

For the present purpose, the central point of the inflation paradigm is that the gravitational potential field prior to the inflationary epoch is expected to have exhibited Gaussian random, spatially uncorrelated fluctuations. These fluctuations at every point, $\mathbf{x}$, in space can be parameterised by the (matter) density perturbation

$$\delta(\mathbf{x}) = \frac{\rho_M|_{\mathbf{x}} - \bar{\rho_M}}{\bar{\rho_M}}, \tag{2.4}$$

where $\rho_M|_{\mathbf{x}}$ is the matter density at the location $\mathbf{x}$. The fluctuation field can be written as a Fourier integral:

$$\delta(\mathbf{x}) = \int_0^\infty \frac{d^3k}{(2\pi)^3}\delta(\mathbf{k})e^{i\mathbf{k}.\mathbf{x}}, \tag{2.5}$$

where $k = 2\pi/\lambda$ is the spatial wavenumber corresponding to a wavelength λ and $\mathbf{k}$ is a wavevector. The Gaussian nature of the fluctuation field implies that its statistics may be wholly characterised by the fluctuation power spectrum, $P(\mathbf{k})$, defined as

$$P(k)\delta_{\rm Dirac}(\mathbf{k} - \mathbf{p}) = \langle\delta(\mathbf{k})\delta^*(\mathbf{p})\rangle \tag{2.6}$$

where the angled brackets indicate averaging over all space. Scale invariance in the dimensionless power spectrum of potential fluctuations in the inflaton scalar field, which is approximately predicted by inflationary models (e.g., Guth & Pi 1982), results (through the Poisson equation) in a primordial matter fluctuation power spectrum of the Harrizon-Zel'dovich-Peebles form (Harrison 1970; Peebles & Yu 1970; Zel'dovich 1972):

$$P_{\rm primordial}(k) \propto k. \tag{2.7}$$

The evolution of $P_{\rm primordial}(k)$ to later times, t, is described by the transfer function, $T(k,t)$, such that

$$P(k,t) = P_{\rm primordial}(k)|T^2(k,t)|. \tag{2.8}$$

The transfer function encodes various physical effects. During the inflation epoch, the exponential growth of the scale factor implied that fluctuation modes with wavelengths $\frac{2\pi}{k} > ct_{\rm exit}$, at a given time $t_{\rm exit}$, exited the horizon and ceased to grow. These modes then re-entered the horizon following inflation as power-law expansion resumed. However, modes that re-entered during the radiation-dominated era could not grow in amplitude, whereas modes re-entering during the matter-dominated era, after the redshift $z_{\rm eq}$ of equality of matter- and radiation-densities, could grow. Hence, the power spectrum at late times retains the primordial form for long-wavelength modes (low k), while the slope is reduced by 4 for short-wavelength modes. The power per logarithmic k-interval given by $k^3P(k,t)$, assuming Equation 2.7 for the primordial power spectrum, is hence proportional to k^4 for long-wavelength modes, and flat for short-wavelength modes.

Fundamental insights into the nature of non-baryonic dark matter were derived by considering the evolution of fluctuations during the matter-dominated era, and the resultant formation of structure in the Universe (e.g., White et al. 1983; Davis et al. 1985). In particular, the hot dark matter model (e.g., Bond & Szalay 1983), which postulates that dark matter consists of relativistic, massive (rest energies $\gtrsim 10\,{\rm eV}$) neutrinos produced in thermal equilibrium with the cosmic background radiation at early times, was ruled out primarily by considering structure formation scenarios (White et al. 1983, 1984). The long free-streaming lengths of neutrinos in the matter-dominated era prior to recombination implied that fluctuations on present-day scales of tens of Mpc and below would be damped (e.g., Bond & Szalay 1983); galaxy clustering at these scales would only be possible

if galaxy formation occurred unreasonably late.

Today, it is generally believed that the matter density of the Universe is dominated by cold dark matter (CDM; Peebles 1982), consisting of unknown massive particles with negligible interaction cross sections and thermal motions. The key initial success of this model was in reproducing the observed distribution of nearby galaxies and galaxy clusters, given the assumption of galaxy formation in high-density peaks of the initial dark matter distribution (Davis et al. 1985), and using the scale-invariant form of the primordial matter power spectrum in Equation 2.7.[2] The next major development was the discovery of anisotropy in the cosmic microwave background (CMB) radiation (Smoot et al. 1992), with the characteristic correlation function expected from scale-invariant primordial scalar fluctuations (Equation 2.7; Bond & Efstathiou 1987). Measurements of the luminosity distances and redshifts of Type 1a supernovae samples by Riess et al. (1998) and Perlmutter et al. (1999) revealed that a cosmological model with a non-zero cosmological constant, Λ was required. The new 'ΛCDM' paradigm was spectacularly confirmed by the first-year results of the Wilkinson Microwave Anisotropy Probe (WMAP; Spergel et al. 2003), which was used to measure the angular power spectrum of the CMB anisotropies over three decades in scale. The detection of baryon acoustic oscillations in the spatial correlation function of low-redshift galaxies (Eisenstein et al. 2005) provided further independent evidence for the ΛCDM model, as did measurements of the large-scale matter distribution through wide-field searches for weak lensing effects (Van Waerbeke et al. 2002).

2.1.2 Structure formation

The theory of galaxy formation in a ΛCDM Universe is based on the cooling and condensation of baryons within the overdensities of dark matter that first decoupled from the Hubble flow (e.g., White & Rees 1978). The subsequent gravitational dynamics of dark matter haloes governs the clustering and merger histories of galaxies. I shall make use of a semi-analytic galaxy formation model (Guo et al. 2011) based on halo merger trees from a numerical dark matter simulation (the Millennium simulation; Springel et al. 2005) in Chapters 5 and 6 of this thesis. Here, I discuss and

[2]Intriguingly, current measurements suggest that $P_{\rm primordial}(k) \propto k^{0.961\pm0.005}$ (Planck Collaboration et al. 2013).

compare standard computational methods for the evolution of the dark matter halo population, and review galaxy formation processes in this context in the following sections (§2.2 and §2.3).

Analytic computations of the evolution of the dark matter halo mass function (Press & Schechter 1974; Bond et al. 1991; Lacey & Cole 1993; Sheth et al. 2001) provide significant insight into cosmological structure over many decades in scale. The evolution of a density perturbation, $\delta(\mathbf{x}, t)$ (Equation 2.4), with redshift is described, for $\delta(\mathbf{x}, z) \ll 1$, according to the linear growth factor $D(z)$:

$$\frac{\delta(\mathbf{x}, z_2)}{\delta(\mathbf{x}, z_1)} = \frac{D(z_1)}{D(z_2)}. \tag{2.9}$$

For the ΛCDM cosmological model, an accurate fit to the growth factor as a function of redshift is given by (Carroll et al. 1992)

$$D(z) = \frac{5\Omega}{2(1+z)} \left(\frac{1}{70} + \frac{209\Omega_M(z)}{140} - \frac{\Omega_M^2(z)}{140} + \Omega_M^{4/7}(z) \right), \tag{2.10}$$

in terms of the redshift-dependent matter density parameter

$$\Omega_M(z) = \Omega_M(0)(1+z)^3[\Omega_\Lambda + \Omega_M(0)(1+z)^3]^{-1}. \tag{2.11}$$

The evolution of initially small overdensities becomes non-linear as the perturbation size approaches unity. It is somewhat straightforward to derive the projected (unphysical) linear overdensity of a halo that has decoupled from the Hubble flow, collapsed, and virialised; this critical value is

$$\delta_c \approx 1.686 \tag{2.12}$$

for a spherical overdensity (e.g., Press & Schechter 1974). The monotonic nature of the ΛCDM growth factor, coupled with the large amount of power in the density-perturbation spectrum, $P(k, t)$, at short fluctuation wavelengths (Equation 2.8 and related discussion), implies that the smallest collapsed dark matter structures formed first, and that larger structures formed mainly through the merging of smaller structures. This is the basis of the so-called 'hierarchical' or 'bottom-up' structure formation scenario.

It is useful to define the variance, $\sigma^2(M, z)$, of the density field at redshift z

smoothed on a mass scale, M, with the spherical top-hat function

$$W_M(r) = \begin{cases} 1, & r \leq R(M) \\ 0, & \text{otherwise} \end{cases} \tag{2.13}$$

where r is the radial coordinate and $R(M) = \left(\frac{3M}{4\pi\rho_M}\right)^{1/3}$. Then, from the Fourier transform of $W_M(r)$ and the convolution theorem,

$$\sigma^2(M,z) = \frac{1}{2\pi^2}\int_0^\infty k^2 P(k,z)\left[\frac{3(\sin(kR) - kR\cos(kR))^2}{(kR)^3}\right]^2 dk, \tag{2.14}$$

where I insert a redshift-dependence into the power spectrum. Now, the probability that a window of radius $R(M)$ has been placed on a *collapsed* halo is given by $\mathrm{erfc}(\delta_c/\sigma(M,z))$, where erfc is the complementary Gaussian error function. Further, the difference in probabilities $\mathrm{erfc}(\delta_c/\sigma(M+dM,z)) - \mathrm{erfc}(\delta_c/\sigma(M,z))$ corresponds to the comoving number density of haloes with masses in the range M to $M+dM$, $\frac{dn}{dM}dM$, multiplied by the comoving volume $\frac{M}{\rho_M}$. Thus, one can derive the Press-Schechter mass function of haloes:

$$\frac{dn}{dM}(M,z) = \sqrt{\frac{2}{\pi}}\left|\frac{d\sigma(M,z)}{dM}\right|\frac{\rho_M}{M}\frac{\delta_c}{\sigma^2(M,z)}\exp\left(\frac{\delta_c^2}{2\sigma^2(M,z)}\right). \tag{2.15}$$

The factor of 2 was inserted with remarkable intuition by Press & Schechter (1974) to account for the contributions to the halo mass function from haloes within underdense regions. However, this factor arose naturally in the identical extended Press-Schechter (EPS) formalism introduced by Bond et al. (1991) for the halo mass function, based on analysing the 'excursion-set' statistics of trajectories of δ as functions of $\sigma(M,0)$ and uniquely identifying redshifts where values of δ at $z = 0$ corresponded to the critical overdensity δ_c. The final iteration of analytic techniques for estimating the halo mass function was the work of 'Sheth-Tormen' formalism (Sheth et al. 2001), which includes a treatment of the ellipsoidal collapse of dark haloes.

The EPS formalism enabled Lacey & Cole (1993) to derive an expression for the

merger rates of individual dark matter haloes:

$$\frac{d^2P}{d\Delta M dt} = \left(\sqrt{\frac{2}{\pi}}\right) \left|\frac{d\delta_c(t)}{dt}\right| \left(\frac{\Delta M}{\sigma_2^2}\right) \left|\frac{d\sigma(M_2,0)}{dM_2}\right| (1-\frac{\sigma_2^2}{\sigma_1^2})^{-3/2} \exp\left(\frac{\delta_c^2}{2(\sigma_2^2-\sigma_1^2)}\right) \tag{2.16}$$

where $\frac{d^2P}{d\Delta M dt}$ is the differential probability of a halo with mass M_1 merging with a halo of mass ΔM to form a halo of mass $M_2 = M_1 + \Delta M$ per unit proper time t per unit ΔM, $\sigma_1 = \sigma(M_1, 0)$, $\sigma_2 = \sigma(M_2, 0)$ and $\delta_c(t) = \delta_c D[z(t)]$. Equation 2.16 can be used along with the halo mass function of Equation 2.15 to construct Monte Carlo realisations of halo merger 'trees', which trace the merger histories of model haloes at $z = 0$ (Somerville & Kolatt 1999). Halo merger trees are the basic component of semi-analytic models of galaxy formation.

The analytic treatment of the dark matter halo population has only been widely applied because of its rough consistency with more computationally-expensive numerical results (Efstathiou et al. 1988). The numerical techniques of smoothed particle hydrodynamics (SPH; Gingold & Monaghan 1977) used, for example, by Davis et al. (1985) to study the non-linear evolution of structure in the dark matter distribution have proved central to our current understanding of the formation of collapsed baryonic objects, such as galaxies and SMBHs. SPH calculations for the evolution of a fluid involve the computation of the positions and velocities of particles, or fluid elements, into which the fluid is split. The force laws between particles are smoothed on a possibly time- and position-variable length scale. For example, Davis et al. (1985) simulated 32^3 dark matter 'particles' smoothed on a scale $L/213$, where L was the side length of the computational box, such that the gravitational interactions between particles separated by less than $L/213$ corresponded to two interpenetrating spheres, and the gravitational interactions between particles separated by greater than $L/213$ corresponded to two point masses. SPH simulations were first used by Efstathiou et al. (1979) to study the mass-distribution of gravitationally-bound CDM clumps, later interpreted as haloes, by identifying clumps of SPH particles using a 'friends-of-friends' algorithm. Navarro et al. (1996) provided a fit to the density profiles of dark matter haloes by re-simulating with higher resolution representative clumps from large-scale CDM simulations; observational tests of such density profiles are part of current investigations into the viability of the CDM model (e.g., Łokas & Mamon 2001).

The Millennium simulation (Springel et al. 2005), which I utilise in this thesis, is likely the most influential numerical study of dark matter ever carried out. The simulation traced 2160^3 particles within a periodic computational box with comoving side length $500\,h^{-1}$Mpc. The adopted concordance ΛCDM cosmology from the first-year WMAP results (Spergel et al. 2003), which I refer to as 'WMAP1', included the matter density parameter $\Omega_M = 0.25$, cosmological constant parameter $\Omega_\Lambda = 0.75$, Hubble constant $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and normalisation of the matter density fluctuation power spectrum $\sigma_8 = 0.9$. The σ_8 parameter is simply the value of $\sigma(M, 0)$ from Equation 2.14 at $z = 0$ for M corresponding to a sphere of radius $8\,h^{-1}$Mpc. With these cosmological parameters, the minimum resolvable halo mass, consisting of 20 particles, was $2.36 \times 10^{10} M_\odot$, and the simulation box side length was 685 Mpc. The dark matter particle distribution was recorded at 63 redshift snapshots, z_i, identified by

$$z_i = 10^{i(i+35)/4200} - 1 \tag{2.17}$$

where the integer i runs from 0 to 62. At each redshift, a friends-of-friends algorithm was used to classify bound dark matter structures. Following Davis et al. (1985), bound particles were classified as having separations of less than 0.2 of the mean particle separation. Approximately 1.8×10^{10} haloes were identified at $z = 0$, corresponding to 49% of the simulation particles. Bound structures within larger haloes were identified as subhaloes using an algorithm described in Springel et al. (2001).

While the best analytic treatments of the dark matter halo population (Sheth et al. 2001) provide a reasonable match to the halo mass function and merger rate predicted by the Millennium simulation, the most commonly used EPS formalism does not (Springel et al. 2005; Fakhouri & Ma 2008). I demonstrate this in Figures 2.2 and 2.3. In Figure 2.2, I show the dark matter halo mass functions at $z = 0$ and $z = 2$ calculated using Equation 2.15 and the WMAP1 cosmological parameters, and compare these with the halo mass functions at these redshifts that calculated from the publicly-available Millennium simulation halo catalogues.[3] The EPS mass functions clearly over-predict the halo counts at all but the very highest masses contained in the Millennium simulation; at the highest masses, the EPS mass functions are below the Millennium results. Similarly egregious results are evident in Figure 2.3, where I compare the merger rates of haloes of different masses at $z = 0.1$

[3]Details are given at http://www.mpa-garching.mpg.de/millennium

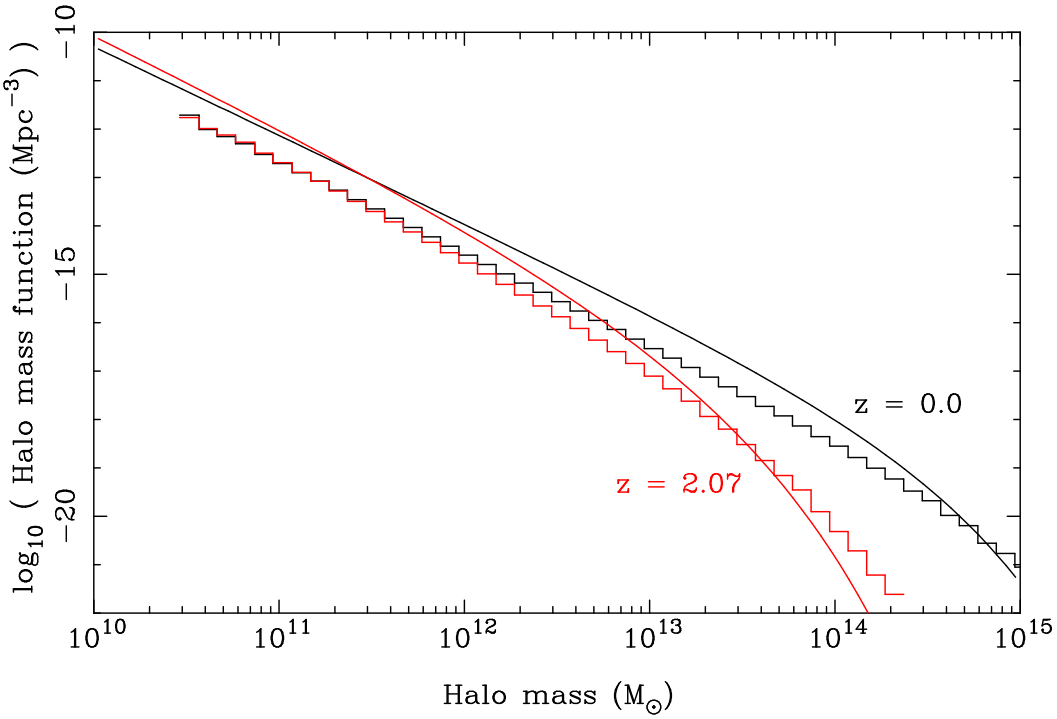

Figure 2.2: A comparison between the dark matter halo mass function calculated using the EPS prediction in Equation 2.15 (smooth curves) and the halo mass function derived from the Millennium simulation (histograms). The black traces denote results for $z = 0$, and the red traces denote results for $z = 2.07$.

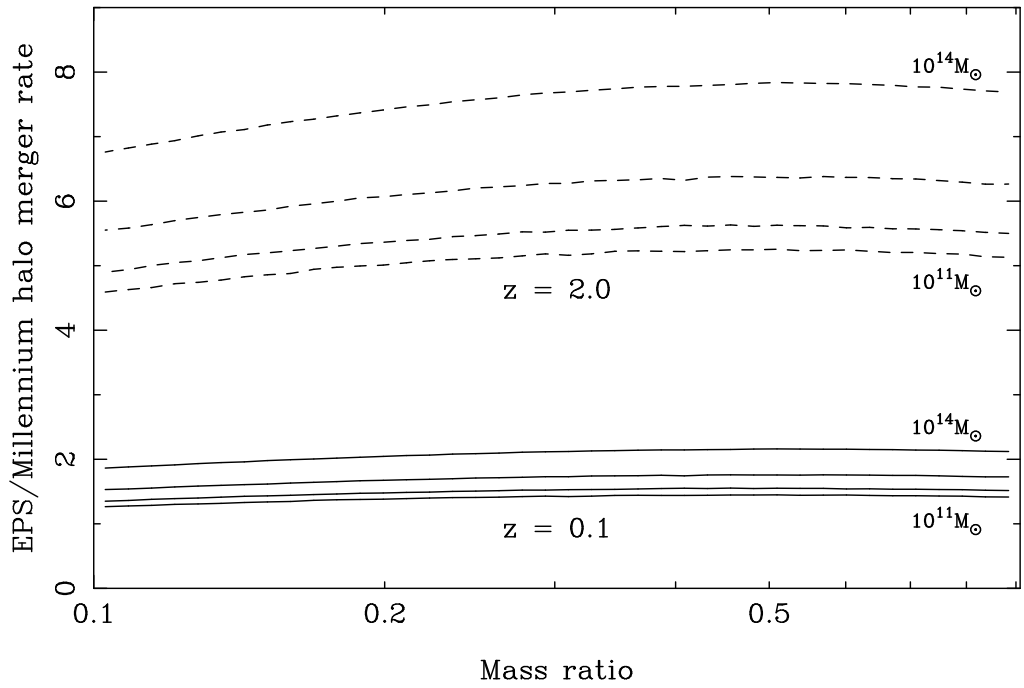

Figure 2.3: A comparison between the rate of mergers onto dark matter haloes calculated using the EPS prediction in Equation 2.16 and using an analytic fit to the Millennium simulation merger rates from Fakhouri & Ma (2008). I consider haloes with masses $10^{11}M_{\odot}$, $10^{12}M_{\odot}$, $10^{13}M_{\odot}$ and $10^{14}M_{\odot}$ at redshifts $z = 0.1$ (solid curves) and $z = 2$ (dashed curves). The curve heights at each redshift increase with increasing halo mass. The results are presented as ratios of the EPS merger rates to the Millennium merger rates for mergers with different mass ratios.

and $z = 2.07$ from the EPS prediction (Equation 2.16) to the analytic fit to the Millennium simulation merger rates from Fakhouri & Ma (2008). The EPS results ubiquitously over-predict the halo merger rate for mass ratios $q = \frac{\Delta M}{M_1}$ between 0.1 and 1/1.1, and the discrepancies increase both with halo mass and redshift.

By construction, numerical simulations of the dark matter halo distribution, such as the Millennium simulation, reliably represent the underlying assumptions of the ΛCDM paradigm. As I have shown, they do so with greater accuracy than analytic approximations such as the EPS formalism. I therefore use the results of the Millennium simulation for the evolution of the dark matter halo distribution in Chapters 5 and 6 of this thesis. However, for numerical dark matter simulations to be used to represent the observed Universe, they must be (*a*) conducted using the correct cosmological parameters, and (*b*) coupled to analytic prescriptions for the formation and evolution of baryonic structures. The WMAP1 parameters, used in the Millennium simulation, have been superseded by successive data releases from the WMAP and *Planck* satellites (e.g., Planck Collaboration et al. 2013), combined with other probes of the large-scale structure of the Universe. Furthermore, large-scale numerical cosmological simulations are, at the time of writing, beginning to concurrently capture both baryon and dark matter structure formation (e.g., the Illustris simulation; Vogelsberger et al. 2014). However, as shall be seen, the effects of updates to cosmological parameters on the conclusions of the thesis are minor, and simulations such as Illustris do not yet fully capture the rarest, most massive SMBHs that are of great importance to my work.

2.2 The formation of galaxies and SMBHs

In this section, I review current understanding of the formation of the first galaxies and SMBHs. This is intended to provide a context for models for the evolution of the galaxy and SMBH populations at later times in the history of the Universe, that I shall discuss subsequently and form the basis for much of this thesis.

Understanding the transformation of the baryon content of the Universe from a neutral, primordial gaseous soup into the exquisite structures observed today is a core problem of astrophysics. Prior to the epoch of recombination at $z \sim 1089$, during which time the observed CMB radiation was emitted, the coupling between photons and baryons damped the gravitational growth of baryon-density fluctua-

tions. During the era immediately following recombination, termed the cosmological dark ages, the Universe was pervaded by neutral gas which evolved under the gravitational influence of the underlying dark matter distribution. The thermal evolution of the gas was decoupled from the CMB after $z \sim 100$, at which redshift the first, Earth-sized dark matter haloes are though to have collapsed and virialised (Diemand et al. 2005). As the gas density and temperature increased within dark matter haloes, the first (Population III) stars were able to form in haloes where sufficient H_2-cooled gas was present to allow fragmentation into Jeans-mass clouds. It is, however, unclear whether the formation of Pop III stars heralded the first galaxies, or in fact suppressed galaxy formation for many tens of Myr. The first galaxies hence consisted either of a combination of Pop III and newer stars, or solely of new generations of stars in heavier dark matter haloes. In either case, the first galaxies were likely to have predominantly contributed to the process of ionising the IGM and enriching it with metals.[4] Reviews of the current understanding of the dark ages and the epoch of re-ionisation (EOR) are given by Barkana & Loeb (2001), Fan et al. (2006) and Bromm & Yoshida (2011).

2.2.1 The first galaxies

Detailed studies of the conditions required to form the first stars (e.g., Abel et al. 2000; Bromm et al. 2002) suggest that they were formed from gas within the largest ($\sim 10^6 M_\odot$) virialised dark matter haloes at $20 \lesssim z \lesssim 30$. The virial temperature of a halo is defined as

$$T_{\rm vir} = \frac{\mu m_p V_c^2}{2k_B}, \tag{2.18}$$

where μ is the mean molecular weight of neutral hydrogen gas, m_p is the proton mass, V_c is the maximum virialised halo circular velocity and k_B is Boltzmann's constant. This quantity represents the virialised gas temperature in a halo in the absence of cooling. The largest collapsed haloes at $20 \lesssim z \lesssim 30$ had $T_{\rm vir} < 10^4\,{\rm K}$, which is the threshold above which cooling dominated by the presence of H_2 is suppressed (Oh & Haiman 2002). These haloes were shown by Abel et al. (2000) and Bromm et al. (2002) to host $\sim 10^3 M_\odot$ of gas at few-hundred Kelvin temperatures,

[4] 'Metals' to astronomers are elements heavier than Helium.

corresponding to the cosmological Jeans mass (Barkana & Loeb 2001) at these redshifts. This gas could then collapse and form $\sim 100 M_{\odot}$ Pop III stars.

However, the fates of the 'minihaloes' that hosted the first stars are uncertain. Negative feedback from the first stars on future star formation, caused by energetic supernovae (Bromm et al. 2003), pressure waves driven by photoionisation heating (e.g., Kitayama et al. 2004), and a H_2-dissociating UV background (Haiman et al. 2000) may have delayed the onset of galaxy formation by up to 10^7 yr (Bromm & Yoshida 2011). However, if the first generation of Population II stars forming after Pop III supernovae in a less pristine IGM had an initial mass function similar to that today, rather than a heavier characteristic mass (Bromm et al. 2009), it is possible that positive stellar feedback effects that enhanced H_2 formation (Ricotti et al. 2001) in fact caused minihaloes to self-enrich with metals and grow to form the first galaxies (Ricotti et al. 2002). In the former, perhaps more widely accepted case (Barkana & Loeb 2001; Bromm & Yoshida 2011), the first galaxies formed at $z \sim 10$ in dark matter haloes hosting baryon masses of $\sim 10^8 M_{\odot}$ corresponding to $T_{\rm vir} > 10^4\,{\rm K}$, which signifies the atomic hydrogen cooling regime. These haloes were capable of re-virialising gas that was heated by previous generations of stars (Wise & Abel 2007) and had sufficiently deep potential wells to retain photoheated gas (Dijkstra et al. 2004). While the gas within these systems is likely to have been turbulent, Wise & Abel (2007) suggest that the gas could cool sufficiently through atomic hydrogen line emission and contributions from heavier elements to support the formation of protogalactic disks.

While the galaxy with the highest confirmed redshift lies at $z = 7.51$ (Finkelstein et al. 2013), nearly 1000 galaxies (Bouwens et al. 2014) have been discovered at $z \gtrsim 7$ using the 'drop-out' technique (Steidel et al. 1996) with the Hubble space telescope (HST). Light emitted by galaxies that is blue-ward of the rest-frame Lyman α wavelength of 1216 Å is absorbed by intervening neutral hydrogen as it is redshifted into Lyman α resonance. Galaxies at redshifts $z > 6$ hence appear only at wavelengths greater than 850 nm, and the wavelength below which galaxies drop out of images increases with the galaxy redshift. The HST Wide Field Camera 3 is used to find candidate galaxies at redshifts up to 10 (e.g., Oesch et al. 2014) by obtaining deep images with various filters and identifying objects that only appear through the reddest filters. Observations (Stark et al. 2013) suggest that the SFR per galaxy rises uniformly with redshift for $z > 5$, while the global SFR den-

sity reduces with redshift (e.g., Oesch et al. 2014). Of the ten candidate galaxies at $z \sim 10$ that are currently known (Oesch et al. 2014), four with detections with the Spitzer space telescope at a wavelength of $4.5\,\mu$m are inferred to have stellar masses of $\sim 10^9 M_\odot$, possibly larger than are predicted by models with the delayed formation of the first galaxies. The measured galaxy luminosity function appears to be consistent with the lower-redshift Schechter function (Schechter 1976) form at all redshifts up to 8, with the most significant evolution occurring at the highest luminosities (Bouwens et al. 2014). However, it is unlikely that these observations probe the full galaxy population at these redshifts. The measured electron scattering optical depth of the CMB (e.g., Planck Collaboration et al. 2013) suggests that if the IGM was re-ionised instantaneously, the EOR occurred at $z = 11.3 \pm 1.1$. A combined analysis of these measurements and galaxy observations by Robertson et al. (2013) showed that an extrapolation of the observed luminosity functions to at least three magnitudes fainter was required, and that the inferred faint galaxy population predominantly contributed to re-ionising the Universe. This is consistent with theoretical models for contributions to the ionising radiation budget at these redshifts (e.g., Duffy et al. 2014).

2.2.2 The first SMBHs

The discovery of quasars with masses $M_\bullet > 10^9 M_\odot$ at $z > 6$ (Fan et al. 2001; Fan 2006; Mortlock et al. 2011) demonstrated that SMBHs with masses comparable to the most massive SMBHs today existed less than $10^9\,$yr after the Big Bang. These rare objects have typically been detected in large optical surveys, such as the Sloan Digital Sky Survey (e.g., Fan et al. 2001), the Canada-France High-z Quasar Survey (Willott et al. 2010) and the United Kingdom Infrared Telescope Deep Sky Survey (Lawrence et al. 2007), and have an inferred comoving space density of $\sim 1\,\mathrm{Gpc}^{-3}$. While the luminosity function of the high-redshift quasars (Willott et al. 2010) is consistent with the observed evolution of the lower-redshift quasar population (Hopkins et al. 2007), little evolution with redshift is observed in intrinsic quasar properties such as spectra and metallicities (Jiang et al. 2007). Deep searches with the Chandra X-ray Observatory for AGN at the locations of drop-out galaxies at $z > 6$, including stacking of the X-ray counts for multiple galaxies, have however not revealed any detections (Willott 2011; Treister et al. 2013). Even as-

suming a 1% fraction of galaxies with AGN and an Eddington ratio of 10%, these results suggest a mean SMBH mass among the searched galaxies of less than $3\times10^7 M_\odot$ if all the galaxies host central SMBHs (Treister et al. 2013). The Eddington ratio, $\lambda_{\rm Edd}$, of an active SMBH is defined as the ratio of the bolometric luminosity, $L_{\rm bol}$, to the Eddington luminosity, $L_{\rm Edd}$, which is given by

$$L_{\rm Edd} = \frac{4\pi G M_\bullet m_p c}{\sigma_T}, \tag{2.19}$$

where σ_T is the Thomson scattering cross-section. The Eddington luminosity corresponds to the limit where the outward radiation pressure balances the gravitational pressure in an isotropic accretion flow, assuming Thomson scattering opacity and a pure hydrogen flow composition.

How could such massive SMBHs be assembled in several hundred Myr? Turner (1991) and Haiman & Loeb (2001) pointed out, in the contexts of quasars detected at redshifts $z > 4$ and $z > 6$ respectively, that the growth of such SMBHs from $\sim 100 M_\odot$ seeds would require accretion at close to the Eddington luminosity for the age of the Universe. This is at odds with the inferred duty cycles of quasars of these masses at $z < 6$, which are expected to be less than 0.5 (Shankar et al. 2013). Thus, it is possible that the early growth of SMBHs was through significantly different mechanisms to SMBH growth at subsequent times.

Two popular mechanisms for the formation of SMBH seeds have emerged in recent years: $\sim 100 M_\odot$ remnants of Pop III stellar supernovae, and more massive objects formed through the direct collapse of cold gas clouds (for reviews, see Volonteri 2010; Volonteri & Bellovary 2012; Haiman 2013). While the popularity of these mechanisms (c.f. Haiman 2013) can be partially ascribed to indirect arguments of theoretical convenience, they do appear to be underpinned by plausible physics, as described below.

Pop III remnants. The compact remnants of the supernovae of Pop III stars, which are expected to form at redshifts $z > 20$, are thought to be a strong function of their masses. For masses between $25 M_\odot$ and $140 M_\odot$, black holes containing around half the stellar mass are expected to form (Zhang et al. 2008), whereas for stars with masses between $140 M_\odot$ and $260 M_\odot$, the pair production instability will cause the entire star to disintegrate (Bond et al. 1984). For masses above $260 M_\odot$, the photodisintegration instability in the stellar

core is expected to accelerate collapse to black holes before the pair instability can act; the black holes again contain approximately half the stellar mass (Fryer et al. 2001). Modelling of the initial mass function of Pop III stars suggests a high mean mass of $\sim 100 M_\odot$ (McKee & Tan 2008; Bromm et al. 2009), although the various feedback processes that set this mass are highly uncertain. Further feedback processes, such as photodissociation of H_2 and metal pollution could in fact suppress Pop III star formation in most haloes with $T_{\rm vir} < 10^4\,{\rm K}$ (e.g., Haiman & Holder 2003), as could residual bulk streaming motions between the IGM and host haloes (e.g., Tanaka & Li 2014).

Direct collapse. In this scenario, gas clouds in dark matter haloes with $T_{\rm vir} > 10^4\,{\rm K}$ are thought to be capable of collapsing without fragmentation into star-forming clumps, and feeding a central supermassive star (SMS; Hoyle & Fowler 1963) with Eddington ratios $\gg 1$. The SMS contracts until it becomes dynamically unstable, and collapses to a black hole if its mass exceeds $10^5 M_\odot$ (Shapiro 2004). If the accretion rate is $\gtrsim 1\,M_\odot\,{\rm yr}^{-1}$, a quasi-star forms as an envelope around the black hole (e.g., Begelman et al. 2006), and the Eddington-limiting accretion rate for the hole corresponds to the limit for the entire system.

Whether or not a gas cloud can collapse in such a manner is an open question. The self-gravitating disks that are expected to form in collapsing gas clouds are extremely susceptible to effective cooling and fragmentation. For example, numerical simulations of gas collapse in $T_{\rm vir} > 10^4\,{\rm K}$ haloes by Shang et al. (2010) showed that some mechanism is required that would prevent the formation of H_2 and hence fragmentation, such as a strong external UV background. The presence of a nearby heavy halo that photodissociates H_2 is expected in some cases (Dijkstra et al. 2008), and sufficiently high Lyman α opacity could also trap the (atomic) cooling radiation (Spaans & Silk 2006). Gas dynamics in the disks could play a role: some low angular momentum gas can be transported to the centre without fragmentation (Lodato & Natarajan 2006), and gravitational instabilities such as the 'bars within bars' mechanism may also fulfil the same function (Begelman & Shlosman 2009). Sethi et al. (2010) show that the tangling of primordial magnetic fields could provide a useful mechanism for heating the gas. However, Omukai et al. (2008) point out that haloes with

$T_{\rm vir} > 10^4\,{\rm K}$ are expected to already be enriched with metals from earlier generations of Pop III stars, and that even with extremely strong external UV backgrounds, a critical metallicity exists above which fragmentation and the formation of star clusters is expected.

Other means of forming SMBH seeds at high redshifts have also been proposed. Star clusters with $\gtrsim 10^6$ members have sufficient negative heat capacity to support runaway core collapse into a single black hole, as long as stars do not evolve off the main sequence on the collapse timescale (Quinlan & Shapiro 1990). Mass segregation may result in less-massive clusters also being prone to collapse (e.g., Begelman & Rees 1978). Primordial black holes may also be formed through extreme density fluctuations in the very early Universe, with masses roughly corresponding to the mass contained within the particle horizon at the time of formation (Carr 2003).

The growth timescales of SMBH seeds are typically defined in terms of Eddington-limited accretion. Let the fraction of the mass accretion flow, $\dot{M}_{\rm acc}$, onto a SMBH that is radiated be ϵ, such that the radiated luminosity is:

$$L_{\rm BH} = \epsilon \dot{M}_{\rm acc} c^2 = \frac{\epsilon \dot{M}_\bullet c^2}{1-\epsilon} = \lambda_{\rm Edd} L_{\rm Edd}, \tag{2.20}$$

where the Eddington luminosity, $L_{\rm Edd}$, and ratio, $\lambda_{\rm Edd}$, are defined in and near Equation 2.19. The ϵ term is known as the radiative efficiency, and is typically set to ~ 0.075 for consistency between the local mass density in SMBHs and the bolometric quasar luminosity function (Chokshi & Turner 1992; Yu & Tremaine 2002; Shankar et al. 2013). Now, given $L_{\rm Edd} = 1.26 \times 10^{31} (M_\bullet / M_\odot)\,{\rm J\,s^{-1}}$, we have

$$t_{\rm growth} = \frac{M_\bullet}{\dot{M}_\bullet} = 450 \frac{\epsilon}{\lambda_{\rm Edd}(1-\epsilon)}\,{\rm Myr} \approx 450\epsilon\,{\rm Myr}, \tag{2.21}$$

where $t_{\rm growth}$ is the e-folding time for the SMBH mass, and the latter approximation assumes Eddington-limited growth (i.e., $\lambda_{\rm Edd} = 1$) and $\epsilon \ll 1$. Following, e.g., Tanaka & Li (2014), the mean Eddington ratio required to form a SMBH with mass $M_\bullet$ from a seed of mass $M_{\rm seed}$ in a time T is given by

$$\lambda_{\rm Edd} = \left(\frac{T}{t_{\rm growth}}\right)^{-1} \ln\left(\frac{M_\bullet}{X_{\rm merge} M_{\rm seed}}\right), \tag{2.22}$$

where $X_{\rm merge}$ is a factor to account for growth by coalescence between SMBHs in dark matter halo mergers. For Pop III seeds, a factor $X_{\rm merge} \approx 30$ is expected, whereas a factor of $X_{\rm merge} \approx 3$ is expected for direct-collapse seeds. To reproduce high-redshift quasar observations, a ratio $\lambda_{\rm Edd} \sim 0.7$ is required for Pop III seeds, and a similar ratio $\lambda_{\rm Edd} \sim 0.6$ is required for direct collapse seeds (Tanaka & Li 2014). Hence, continuous, near Eddington-limited accretion is required to grow SMBHs at redshifts $z > 6$.

Various authors have attempted to explain such vigorous growth. One class of models (e.g., Volonteri & Rees 2005) posits shorter phases of super-Eddington accretion, possibly enabled by the trapping of photons in an accretion flow that significantly exceeds the Eddington-limiting rate (Begelman 1978; Wyithe & Loeb 2012). However, unless H_2 cooling can be completely avoided, such accretion rates cannot be sustained (Shang et al. 2010), and in actuality only a small fraction of the infalling mass is expected to reach the SMBH horizon and avoid being funnelled into an outflow (e.g., Blandford & Begelman 1999). Other, similar instabilities could result in the limiting mass-accretion rate being exceeded by a factor of $10-100$ (Begelman 2002). The second class of model (e.g., Tanaka & Haiman 2009; Tanaka 2014) simply attempts to explain the high-redshift quasars in terms of conventional ideas for SMBH growth through mergers and sub-Eddington accretion, possibly also triggered by halo-halo mergers (Tanaka 2014). The rapid merging of high-redshift haloes implies that merger-triggered accretion timescales corresponding to halo dynamical times are sufficient (Tanaka 2014). These models, however, make somewhat optimistic assumptions about the ubiquity of accretion and the occupation fraction of SMBHs in haloes (Haiman 2013). Further, neither class of model is consistent with the apparent lack of lower-mass AGN at $z > 6$ (Treister et al. 2013).

2.3 Co-evolution of SMBHs and galaxies

Understanding the fascinating connections between the growth processes of galaxies and SMBHs is a central scientific motivation for this thesis. There is often discordance between theoretical studies concerning the high-redshift Universe (e.g., Tescari et al. 2014) and the lower-redshift Universe (e.g., Vogelsberger et al. 2014); models typically attempt to reproduce observables in these different regimes independently. The reasons for this (e.g., Benson 2010) are at least three-fold:

- higher-redshift observables tend to represent only the rarest, brightest objects, that require large-volume simulations which are computationally expensive to extend to lower redshifts with sufficient resolution;

- models are typically tuned to, and tested against, observations only at the redshifts of interest; and

- observations at lower redshifts are significantly richer than those at higher redshifts; hence, models of differing complexity are currently motived for different redshift regimes.

A self-consistent simulation that reproduces important properties of galaxies and SMBHs from our earliest to latest observations does not currently exist, although some studies (e.g., Wyithe & Loeb 2003b; Fanidakis et al. 2013) come close. While the models used in this thesis are restricted in scope to self-consistently reproducing lower-redshift observables alone, it is important to acknowledge that they provide partial pictures of the history of the Universe, even as it is currently known.

2.3.1 Correlations between SMBH masses and galaxy properties

Correlations have been identified between the masses of SMBHs and various properties of galaxy spheroids, or bulges, in the local Universe (Kormendy & Ho 2013). It is important to recall, however, that correlations only provide indirect evidence for causal connections. Based on some of the earliest attempts at extragalactic SMBH mass measurements, in the Andromeda spiral galaxy M31 and the dwarf elliptical M32, Dressler & Richstone (1988) pointed out that the ratio of the SMBH masses was significantly closer to the ratio of the spheroid luminosities, $L_{\rm bul}$, than to the ratio of the total luminosities. The review on SMBH mass measurements by Kormendy & Richstone (1995) demonstrated log-linear correlations between SMBH mass and $L_{\rm bul}$ using seven objects. Magorrian et al. (1998), using dynamical modelling of photometry and spectrometry of a large sample of 36 nearby galaxies, confirmed the $M_\bullet - L_{\rm bul}$ relation, and also derived an approximately constant ratio $M_\bullet/M_{\rm bul} \sim 0.006$, where $M_{\rm bul}$ is the bulge mass. This is perhaps expected from the elliptical galaxy fundamental plane relation between the mass-to-light ratio and

$L_{\rm bul}$ (Faber et al. 1987). However, the $M_\bullet - L_{\rm bul}$ and $M_\bullet - M_{\rm bul}$ relations were apparently superseded by the discovery of a tighter (lower inferred intrinsic scatter) log-linear correlation between SMBH masses and the spheroid velocity dispersions, σ_c (Ferrarese & Merritt 2000; Gebhardt et al. 2000). Indeed, based on samples of 12 and 26 galaxies respectively, these early papers suggested that the scatter in the measured correlations was dominated by measurement errors alone. Soon afterwards, a correlation between $M_\bullet$ and the masses of host dark matter haloes was reported by Ferrarese (2002), who found an empirical relation between σ_c and the galaxy circular velocity, V_c, at radii where the rotation curves were flat.

While these results generated a fantastic amount of theoretical interest, further observational results present a significantly revised picture (Kormendy & Ho 2013). Early SMBH mass measurements (e.g., Magorrian et al. 1998) were shown to be systematically wrong in some cases (e.g., McConnell et al. 2011) and biased high because of the use of ground-based photometric data with poor spatial resolution that did not properly resolve the SMBH sphere of gravitational influence (Ferrarese & Merritt 2000; Gebhardt et al. 2003). On the other hand, the inclusion of the dynamical effects of dark matter in models for SMBH and bulge masses was shown to increase $M_\bullet$ estimates and reduce $M_{\rm bul}$ estimates (Gebhardt & Thomas 2009). Switching from visible to K-band photometry led to less self-absorption and contamination from young disk stars in spiral galaxy systems (Marconi & Hunt 2003) and hence more accurate mass estimates. SMBH masses measured using analyses of ionised gas rotation curves have also been shown to be systematically biased low (Kormendy & Ho 2013). The definition of σ_c is also important: most results for the $M_\bullet - \sigma_c$ relation calculate σ_c for entire galaxy bulges, whereas excluding regions of bulges within the spheres of influence of SMBHs can significantly reduce σ_c estimates in more massive systems (McConnell & Ma 2013). Some authors have calculated $M_{\rm bul}$ using bulge mass-to-light ratios derived either through correlations from the fundamental plane of elliptical galaxies (McLure & Dunlop 2002) or through the virial theorem (e.g., Marconi & Hunt 2003); these techniques are not consistent with more precise estimates of $M_{\rm bul}$ made using dynamical models (e.g., Häring & Rix 2004). The relation between σ_c and V_c was also found not to hold when large galaxy samples were considered, debunking any fundamental correlation between SMBH and halo

mass (Ho 2007; Kormendy & Bender 2011).[5] Finally, the recognition that SMBH masses do not correlate with properties of the central regions of spiral galaxies with more disk-like bulges, or 'pseudobulges' (Kormendy et al. 2011), has led to a selection of spiral systems being removed from fits for SMBH-spheroid relations (McConnell & Ma 2013; Kormendy & Ho 2013).

Currently, it appears that in the local Universe *the only extant correlations between SMBHs and bulk galaxy properties are between SMBH masses and the masses and velocity dispersions of bulges.* Bulge luminosities, $L_{\rm bul}$, may be comfortably taken as proxies for bulge masses, and it also appears that the intrinsic scatters in the $M_\bullet - M_{\rm bul}$ and $M_\bullet - \sigma_c$ relations are the same (Kormendy & Ho 2013). Furthermore, it must be emphasised that bulges are found in elliptical and large spiral galaxies alone, and that SMBHs do not correlate with galaxy disks. From Kormendy & Ho (2013), the currently-determined $M_\bullet - M_{\rm bul}$ relation is given by

$$\frac{M_\bullet}{10^9 M_\odot} = \left(0.49^{+0.06}_{-0.05}\right) \left(\frac{M_{\rm bul}}{10^{11} M_{\rm odot}}\right)^{1.16\pm0.08}, \tag{2.23}$$

and the $M_\bullet - \sigma_c$ relation is given by

$$\frac{M_\bullet}{10^9 M_\odot} = \left(0.309^{+0.037}_{-0.033}\right) \left(\frac{\sigma_c}{200\,{\rm km\,s^{-1}}}\right)^{4.38\pm0.29}. \tag{2.24}$$

I quote 68%, or 1 sigma, confidence intervals for the estimated parameters. Recently, following the discovery of numerous SMBHs with masses $M_\bullet > 10^9 M_\odot$ (e.g., Gebhardt & Thomas 2009; McConnell et al. 2011; Rusli et al. 2013), it has been noted that the $M_\bullet - \sigma_c$ relation appears to saturate at the high-$M_{\rm bul}$ end, taking on a steeper form (McConnell & Ma 2013; Kormendy & Ho 2013). These galaxies correspond to massive ellipticals with 'core'-like centres, corresponding to a central intensity depletion relative to the more standard cusp-like density profiles. Indeed, the SMBH masses in these systems appears to strongly correlate with the inferred core mass deficits (Graham 2012; Kormendy & Ho 2013). A different, broken power law form for the $M_\bullet - M_{\rm bul}$ relation was reported by Scott et al. (2013), which emphasises the differences between galaxies with cores at high masses and cusps at low masses.

[5]This is not to say that a correlation between SMBH masses and halo masses does not exist. However, this correlation is explicable simply by the fact that bigger galaxies live in bigger haloes.

SMBH mass measurements beyond the local Universe using AGN properties are exceedingly uncertain, and estimates of redshift-evolution in correlations between SMBHs and bulges are affected by a number of biases (Lauer et al. 2007; Shen & Kelly 2010). However, for the sample of actively accreting SMBHs with mass estimates, along with associated bulge mass estimates, typically higher $M_\bullet/M_{\rm bul}$ ratios are observed at higher redshifts, for example by a factor of $2-3$ at $z=2$ (Kormendy & Ho 2013). A larger-than-expected scatter in the measured $M_\bullet/M_{\rm bul}$ ratios is also observed, although various unknown systematic effects may also be present. In contrast to these results, however, the strong similarity between the cosmic SFR density and the SMBH growth rate density as a function of redshift (Madau & Dickinson 2014) is striking. Both rates peak at redshifts $z\sim 2$, and fall away at higher and lower redshifts. A ratio of 0.001 is found at all redshifts between the volume-averaged rates of stellar mass formation and SMBH growth (Heckman et al. 2004).

2.3.2 Galaxy and SMBH co-evolution mechanisms

Most suggested mechanisms for the co-evolution of SMBHs and galaxy bulges are built on the concept of AGN feedback (Fabian 2012).[6] Furthermore, AGN feedback forms an important component of models which successfully predict properties of the galaxy and SMBH population, such as the one I use in Chapters 5 and 6 of this thesis (Guo et al. 2011).

While AGN are *expected* to affect their surrounds through some combination of radiation-driven winds (Silk & Rees 1998) and the kinetic / mechanical action of relativistic particle jets (e.g, Peterson & Fabian 2006), the direct observational motivation for these effects remains somewhat tenuous. Powerful outflows at velocities of $\sim 0.1c$ from quasars accreting with Eddington ratios $\lambda_{\rm Edd}\sim 0.1$ are inferred from X-ray absorption lines that are blue-ward of the systemic velocity (Pounds 2013). These are interpreted as radiation-driven winds and have possible mass-outflow rates of up to 10% of the accretion rates (Fabian 2012). In ultraluminous infrared galaxies (ULIRGs), which are thought to be canonical examples of recent gas-rich galaxy mergers triggering intense star formation activity and

[6] For an alternative model, see Peng (2007) for a discussion of SMBH-bulge correlations being generated through the statistics of numerous mergers.

AGN (Sanders & Mirabel 1996), energetic molecular galactic-scale outflows are commonly observed (e.g., Spoon et al. 2013), although the link between these outflows and AGN is unclear. Evidence for kinetic feedback from AGN with Eddington ratios $\lambda_{\rm Edd} \lesssim 0.01$ is stronger. In massive galaxy clusters where the gas cooling time is less than the infall time, the production of cold gas through 'cooling flows' should result in more star formation activity than is observed (Rees & Ostriker 1977; Fabian 1994). AGN feedback offers a means for quenching cooling flows, as evidenced by observations of bubbles in the hot-gas distributions of many clusters, which are 'filled' by radio jets powered by SMBHs (e.g., Fabian et al. 2008; McNamara & Nulsen 2012).

However, the theoretical importance of AGN feedback cannot be overstated. The analytic derivations by Silk & Rees (1998) and Fabian (1999) of $M_\bullet \propto \sigma_c^5$ and $M_\bullet \propto \sigma_c^4$ respectively, both of which are quite close to the current value of 4.38 ± 0.29 (Equation 2.24), was a particular motivating factor. To illustrate the argument of Silk & Rees (1998), consider a galaxy as an isothermal sphere of radius r, which may have formed as a result of the major merger between two gas-rich disk galaxies (e.g., Benson 2010) that triggered the efficient cold gas accretion onto a central SMBH. The galaxy mass is then

$$M_{\rm gal} = \frac{2\sigma_c^2 r}{G}, \tag{2.25}$$

where I assume the velocity dispersion to be independent of the SMBH. Given that the galaxy is in hydrostatic equilibrium, the free fall time is simply r/σ_c, and the maximum gas collapse rate is

$$\dot{M}_{\rm collapse} = \frac{f_{\rm gas} M_{\rm gal} \sigma_c}{r} = \frac{2 f_{\rm gas} \sigma_c^3}{G}, \tag{2.26}$$

where $f_{\rm gas}$ is the gas fraction of the galaxy. This corresponds to a power of order $P_{\rm collapse} \sim \frac{f_{\rm gas}\sigma_c^5}{G}$, which can be equated to the quasar luminosity $\lambda_{\rm Edd} L_{\rm Edd}$ such that

$$M_\bullet \sim \frac{f_{\rm gas}\sigma_c^5 \sigma_T}{4\pi \lambda_{\rm Edd} G^2 m_p c}. \tag{2.27}$$

This can be viewed as a limiting mass for SMBHs, if energy balance between quasar radiation and the inflowing material is a sufficient condition for equilibrium. If the quasar radiation power exceeds the inflow power, the flow would be quenched and

the galaxy evacuated of cold gas. However, Fabian (1999) argued that momentum balance is a better condition, and derived

$$M_\bullet \sim \frac{f_{\rm gas}\sigma_c^4\sigma_T}{\pi\lambda_{\rm Edd}G^2m_p}. \tag{2.28}$$

Both Equations 2.27 and 2.28 encapsulate the concept of the self-regulated growth of SMBHs in luminous quasar phases through radiation-pressure feedback.

As hinted at above, major[7] mergers between gas-rich spiral galaxies, with gas fractions $f_{\rm gas} \gtrsim 0.25$ are thought to be crucial events in the growth histories of SMBHs and bulges (Benson 2010). The consistency between the observed bolometric quasar luminosity function and the local mass density in SMBHs (Soltan 1982; Chokshi & Turner 1992; Yu & Tremaine 2002; Shankar et al. 2013) suggests that the bulk of SMBH growth did indeed occur through radiatively-efficient accretion with Eddington ratios $\lambda_{\rm Edd} \sim 0.1$. These quasar phases are easily triggered through major mergers in numerical SPH simulations (e.g., Robertson et al. 2006), where it is typically assumed that that the central SMBHs in the merging galaxy pair quickly coalesce prior to accretion commencing (although see Van Wassenhove et al. 2012). In such simulations, mergers also give rise to extreme bursts of star formation and supernova-wind-driven feedback (e.g., Bournaud et al. 2005; Robertson et al. 2006; Cox et al. 2008) which, together with the SMBH feeding and subsequent radiative feedback, leaves behind spheroidal galaxies without any cold gas. Minor mergers have been shown to result in the formation of combined spheroids and disks (Bournaud et al. 2005).

2.3.3 Cosmological-scale galaxy and SMBH evolution models

Cosmological-scale models that investigate major-merger-triggered, self-regulated bulge and SMBH growth in the ΛCDM framework have been successful in reproducing the local SMBH-bulge correlations, the quasar luminosity function, as well as the local galaxy luminosity function and the cosmic SFR history. Haehnelt et al. (1998) attempted to match the observed luminosity functions of galaxies and quasars at $z = 3$ with the EPS dark matter halo mass function, assuming a selection of cosmological models. They showed that while the galaxy

[7]Differing definitions for 'major' exist, although typical mass ratios are greater than 0.25.

luminosity function was matched by the halo mass function assuming a constant halo mass to galaxy light ratio over less than three magnitudes in scale (contrary to findings beyond their magnitude range), matching the quasar luminosity function was more complex. A scaling of $M_\bullet \propto V_{\rm vir}^5 \propto M_{\rm halo}^{5/3}$ was required if quasar lifetimes comparable to the Salpeter (1964) e-folding time were assumed for reasonable radiative efficiencies and $\lambda_{\rm Edd} = 0.1$. Different scalings of $M_\bullet$ with the halo mass $M_{\rm halo}$, for example a linear trend suggested by Haiman & Loeb (1998), required unreasonably short quasar lifetimes.

Wyithe & Loeb (2003b) describe a remarkably prescient model for the quasar luminosity function and the local SMBH-bulge relations based on the assumption that quasars stop growing when the radiated energy exceeds the binding energy of the surrounding gas after one dynamical time. This implies that quasar lifetimes were set by the gas dynamical times. The only free parameter, fixed using the inferred local mass density in SMBHs, was the fraction of the radiated luminosity that quasars on average return to their surrounds, multiplied by the Eddington ratio. The model reproduced the relation observed at the time between SMBH masses and dark matter halo masses (Ferrarese 2002). By assuming that mechanical feedback alone from supernova-driven winds set a maximum stellar mass per dark matter halo, the linear $M_\bullet - M_{\rm bul}$ relation was also reproduced. Predictions of the model for which observational evidence has since been gathered were the existence of $M_\bullet \sim 3 \times 10^{10} M_\odot$ SMBHs at the centres of galaxy clusters (McConnell et al. 2011) and an increasing $M_\bullet / M_{\rm bul}$ ratio with redshift (e.g., Kormendy & Ho 2013).

Kauffmann & Haehnelt (2000) presented an extensive semi-analytic model for the development of the local galaxy stellar mass function (GSMF) and the quasar luminosity function by employing Monte Carlo realisations of dark matter halo merger trees combined with analytic prescriptions for galaxy and SMBH evolution. This model forms the basis for that used in Chapters 5 and 6. At early times, each halo was assigned a cold gas disk that was a constant fraction of its mass, which formed stars according to the Kennicutt (1998) law multiplied by a redshift-dependent term. Some gas was heated by supernova feedback (Kauffmann & Charlot 1998), which was re-cooled using the cooling functions of White & Frenk (1991). These cooling flows were disabled ad-hoc in particularly massive haloes to shut off late-time star formation and hence match the local GSMF. Major mergers, with mass ratios > 0.3,

resulted in a gas mass

$$M_{\rm acc} = \frac{f_{\rm BH} M_{\rm cold}}{1 + (280\,{\rm km\,s^{-1}})^2 / V_c^2} \tag{2.29}$$

being accreted onto a central SMBH formed from the coalescence of the two previously-existing SMBHs. If no SMBHs were present in the merging galaxies, a new SMBH with a mass $M_{\rm acc}$ was assumed to be created in the merger. Here, $M_{\rm cold}$ is the total cold gas available; the remaining cold gas (with mass $M_{\rm cold} - M_{\rm acc}$) was prescribed to form stars. The stars of the two merging galaxies, along with the new stars, were added to the bulge of the new galaxy. In minor mergers, the stars and cold gas of the satellite galaxies were added to the disks of the bigger galaxies. The metal enrichment of galaxies and galaxy ejecta was also modelled (Kauffmann & Charlot 1998), and a metallicity-dependent stellar population model was used (Bruzual & Charlot 2003). The key findings of Kauffmann & Haehnelt (2000) were that, assuming a ΛCDM cosmological model, the GSMF at $z \sim 0$, the evolution of the cosmic SFR and cold gas densities, and the quasar luminosity function up to $z \sim 2$ could all be self-consistently matched by assuming that SMBHs grew entirely through accretion in major mergers which also produced elliptical galaxies, and that the star formation efficiency decreased with redshift. While the self-regulation of SMBH growth was not explicitly modelled by Kauffmann & Haehnelt (2000), AGN feedback is a plausible physical mechanism for setting the value of $f_{\rm BH}$ in Equation 2.29.

Numerous iterations (Guo et al. 2011, and references therein) of the semi-analytic model of Kauffmann & Haehnelt (2000) have since been implemented within the halo merger trees from the Millennium simulation (Springel et al. 2005). Of particular note here are the studies of Croton et al. (2006) and Marulli et al. (2008). Croton et al. (2006) solved the problem of the necessity for the cut-off of cooling flows in massive dark matter haloes by introducing a prescription for 'radio-mode' feedback from AGN. In particular, they assumed SMBHs to be continually accreting hot gas at the rate

$$\dot{M}_\bullet = \kappa_{\rm AGN} \frac{M_\bullet}{10^8 M_\odot} \frac{f_{\rm hot}}{0.1} \left(\frac{V_{\rm vir}}{200\,{\rm km\,s^{-1}}} \right)^3 , \tag{2.30}$$

where $f_{\rm hot}$ is the fraction of the halo mass in hot gas, and the parameter $\kappa_{\rm AGN} = 6 \times 10^{-6} M_\odot\,{\rm yr}^{-1}$ was tuned such that the high end of the GSMF was accurately

reproduced. The accretion was in turn prescribed to inject energy through kinetic feedback (Begelman 2003; McNamara & Nulsen 2012) into the hot-gas reservoirs of massive galaxies, hence quenching cooling flows and late-time star formation. Marulli et al. (2008) investigated the match between a slight modification of the Croton et al. (2006) model and observations of the SMBH population. They found that the intrinsic scatters and forms of the $M_\bullet - L_{\rm bul}$ and $M_\bullet - M_{\rm bul}$ relations were reproduced, along with the $M_\bullet - \sigma_c$ relation given an assumed $V_c - \sigma_c$ relation (Baes et al. 2003). By inserting a mild redshift-dependence into the factor $f_{\rm BH}$ in Equation 2.29, Marulli et al. (2008) were also able to reproduce the quasar luminosity function for $z < 6$, as well as the inferred local SMBH mass function.

Hydrodynamical numerical simulations on cosmologically significant scales which follow both dark matter and baryons have developed significantly since the early work of Katz et al. (1992), who simulated 32^3 dark matter particles and 32^3 baryonic gas particles within a comoving box of side length 22 Mpc. The state-of-the-art today includes the Illustris project (Vogelsberger et al. 2014) and the MassiveBlack II simulation (Khandai et al. 2014), which include 2×1792^3 and 2×1820^3 particles respectively within boxes of side lengths > 100 Mpc. In principle, such simulations provide the most accurate representations of theoretical predictions for the galaxy and SMBH population, because they directly solve for fundamental interactions between matter resolution elements. However, the limited resolutions of the most commonly used SPH techniques require 'subgrid' analytic prescriptions for the physics of star formation, SMBH accretion, feedback processes and cooling (e.g., Khandai et al. 2014). This problem, along with numerical artefact issues, is partially solved in the most recent 'moving mesh' codes (Springel 2010). In the $(106.5\,{\rm Mpc})^3$ Illustris simulation, Vogelsberger et al. (2014) used a moving mesh approach to successfully, for the first time, directly reproduce the *morphological* properties of the observed population of galaxies at redshifts up to $z \sim 6$, including their chemical compositions. The Illustris simulation also successfully reproduces observed properties of SMBHs, and requires AGN feedback that is either radiative or kinetic depending on the accretion rate. While semi-analytic models have been immensely successful in efficiently reproducing bulk properties of the observed galaxy and SMBH populations, direct hydrodynamical simulations are beginning to approach semi-analytic models in scale and provide a significantly more detailed theoretical view of the Universe.

It is important to recognise that neither semi-analytic models nor cosmological hydrodynamical simulations, both of which provide statistically rich samples of galaxies and SMBHs, actually provide accurate representations of comoving volumes within our Universe. Instead, they are most useful as consistency checks of physically-motivated prescriptions for dark matter and baryons, be they as fundamental as basic gravitation and hydrodynamics or more approximate prescriptions for the evolution of bulk properties. In this respect, the number of free parameters is less important than the number of significant physical effects included in the models, and the values of free parameters are typically hard to directly 'measure' through observations. A second use of these models is as a guide for what future observations may see, as I shall demonstrate in this thesis. However, while the models are predictive in this sense, it is typically hard to use these predictions to falsify them. Instead, albeit written with a touch of cynicism, altering the free parameters or adding new physics can usually resolve inconsistencies between the models and data.

2.4 Some missing pieces

Semi-analytic models implemented in numerical dark matter simulations such as the Millennium simulation (Springel et al. 2005; Guo et al. 2011), as well as fully numerical simulations such as Illustris (Vogelsberger et al. 2014; Genel et al. 2014), provide a fantastic match to numerous observables of the populations of galaxies and SMBHs across cosmic time. They do not, however, comprise fully constrained physics. Underlying this thesis are the broad problems of characterising SMBH growth and the cosmological rate of SMBH-SMBH coalescences. As shall be discussed in Chapter 3, this is accomplished by searching for gravitational waves (GWs) from binary SMBHs formed in galaxy mergers. In this subsection, I outline the importance of new observational information on the mass function of SMBHs at various redshifts, as well as on the rate of binary SMBH coalescences.

Semi-empirical estimates of the mass function of SMBHs, $n(M_\bullet, z)$, at various redshifts (e.g., Shankar et al. 2009, 2013) rely on solving the following continuity

equation (Small & Blandford 1992):

$$\frac{\partial n(M_\bullet, z)}{\partial z} = -\frac{\partial \langle t_{\rm growth}^{-1} \rangle M_\bullet n(M_\bullet, z)}{\partial M_\bullet} \frac{dt}{dz} \tag{2.31}$$

where $\langle t_{\rm growth}^{-1} \rangle$ is the mean SMBH growth timescale defined in Equation 2.21, and $dt/dz = [(1+z)H(z)]^{-1}$. Equation 2.31 is typically solved by assuming an initial $n(M_\bullet, z)$ and numerically integrating towards lower redshifts. The quasar luminosity function is employed by relating $M_\bullet$ to a specific quasar luminosity by assuming a redshift-independent radiative efficiency ϵ, which is tuned to match the independently-estimated local SMBH mass function. The growth rate, $\langle t_{\rm growth}^{-1} \rangle$, may depend on redshift and the properties of the SMBH population at a given redshift. For example, Shankar et al. (2009) found marginal evidence for a lengthening of $t_{\rm growth}$ for massive, low-redshift SMBHs, with corresponding ramifications for the evolution of the SMBH population.

However, such estimates of the SMBH mass function are at best indirect, and do not reveal anything significantly different from the more complex models discussed above that also reproduce observations of the local SMBH population and the quasar luminosity function. Hence, there is a clear motivation for a new observable which directly depends on SMBH masses beyond the local Universe. Such an observable would provide independent constraints on the growth rate of SMBH at various redshifts, allowing for a more nuanced investigation of the redshift-evolution of radiative efficiencies and Eddington ratios. Current freedoms in tuning the efficiency of radiative and kinetic feedback from AGN would also be diminished.

A second pair of important quantities in understanding the evolution of the galaxy and SMBH populations is the rate of galaxy mergers and SMBH-SMBH coalescences; neither are well constrained by observations (e.g., Conselice 2014; Colpi 2014). Theoretical predictions for the major merger rate of galaxies with stellar masses $M_* > 10^{10} M_\odot$ vary by more than an order of magnitude (Hopkins et al. 2010), while observational estimates today have at least 0.5 dex uncertainty for $z < 1$ and significantly greater uncertainties at higher redshifts. The theoretical uncertainty stems in particular from the various definitions of galaxy merger timescales (c.f. Hopkins et al. 2010; McWilliams et al. 2014), and the observational uncertainty is caused by the necessity for detailed morphological information to identify galaxies in the process of merging (c.f. Conselice 2014). Observational

errors in the measured GSMF also contribute to uncertainty in the merger rates of galaxies of specific masses.

The situation is worse for the coalescence rate of pairs of SMBHs. Numerous searches have been conducted for dual AGN and gravitationally-bound binary SMBHs (Burke-Spolaor 2013; Bogdanovic 2014; Colpi 2014). These searches have resulted in tens of secure detections of AGN with separations between seven and several hundred parsecs within individual galaxies, along with circumstantial evidence for binary SMBHs at the centres of a few galaxies (e.g., Valtonen et al. 2008; Eracleous et al. 2012). However, an empirical estimate of the SMBH-SMBH coalescence rate has not yet been achieved.

One particular reason for interest in the rate of galaxy mergers at low redshifts is the observed strong decrease in galaxy sizes at fixed M_* with redshift (Conselice 2014, and references therein). While this is observed for both early- and late-type galaxies, the effect is more pronounced for early-type spheroidal systems. There is significant debate as to the relative contributions to the process of growing galaxies of major and minor galaxy mergers (e.g., López-Sanjuan et al. 2012), changes in the formation processes of galaxies with increasing redshift caused by increasing cold-gas fractions (Khochfar & Silk 2006) and the 'puffing up' of galaxies due to galactic winds and stellar evolution processes (Fan et al. 2008).

As discussed above, major mergers of disk galaxies are thought to form massive elliptical galaxies, and this merger rate hence corresponds to the formation rate of ellipticals. Such mergers are also thought to trigger accretion onto SMBHs in high-Eddington-ratio quasar phases. Hence, constraints on the major merger rate are directly relevant to significant events in models for galaxy formation. An observable that (additionally) reflects the coalescence rate of pairs of SMBHs would provide highly useful information on the growth of SMBHs and the relationship between SMBHs and host galaxies. For example, the currently-unconstrained occupation fraction of SMBHs in galaxies at all redshifts may be less than unity, either because of high-velocity kicks imparted to SMBHs formed from binary coalescence (e.g., Gerosa & Sesana 2014) or as a relic of rare SMBH formation processes (e.g., Volonteri & Bellovary 2012). Mass accretion onto SMBHs in major galaxy mergers may also occur both before and after SMBH-SMBH coalescence (e.g., Van Wassenhove et al. 2012), contrary to the standard assumption of accretion following coalescence. Finally, the mass-deficits which are observed

at the centres of most massive elliptical galaxies relative to cuspy density profiles (e.g., Faber et al. 1997) are commonly ascribed to the 'scouring' of stars by binary SMBHs, which formed during the mergers that also formed the galaxies (Milosavljević & Merritt 2001).

Chapter 3

Binary black holes, gravitational waves and pulsar timing arrays

I review important physics and physical methods underlying this thesis. First, I discuss the body of primarily theoretical studies which show that the formation of a binary supermassive black hole (SMBH) is expected upon the merger of two galaxies. I then introduce fundamental concepts of gravitational waves (GWs) from binary systems, and also present a formal, partially original description of a stochastic, isotropic GW background, such as may be created by the summed signals from all binary SMBHs. Finally, I present an overview of pulsar timing techniques and the effects of GWs on timing measurements.

3.1 Gravitational waves from binary SMBHs

In this section, I describe the current theoretical understanding of the formation of binary SMBHs in galaxy mergers, and quantify the gravitational radiation emitted by these systems. This sets the stage for the basic methodology of this thesis, which is to model the gravitational radiation from the cosmological population of binary SMBHs, and to test these models using long-term pulsar timing observations.

While much of this theory is observationally motivated, with, for example, ever more realistic galaxy models being considered, there is no more than circumstantial evidence of the existence of binary SMBHs. A binary SMBH, consisting of SMBHs with masses $M_2 \leq M_1$, is typically defined as forming when the stellar (and gaseous)

mass within the orbit of the lighter SMBH is less than or equal to M_1. If a cuspy stellar density profile is assumed within the gravitational influence radius, $r_{\rm inf}$, of the heavier SMBH (Bahcall & Wolf 1976), and a profile corresponding to an isothermal sphere is assumed beyond $r_{\rm inf}$, (Sesana 2010):

$$\rho_*(r) = \begin{cases} \rho_{\rm inf}\left(\frac{r}{r_{\rm inf}}\right)^{-\gamma}, & r \leq r_{\rm inf} \\ \frac{\sigma_c^2}{2\pi G r^2}, & r > r_{\rm inf} \end{cases} \tag{3.1}$$

where $\rho_*(r)$ is the stellar density at radius r, $\rho_{\rm inf} = \rho_*(r_{\rm inf})$, and γ is the cusp power law index. By equating $\rho_{\rm inf}$ with the stellar density at $r_{\rm inf}$ predicted by the isothermal sphere model,

$$r_{\rm inf} = (3-\gamma)\frac{GM_1}{\sigma_2}. \tag{3.2}$$

From the $M_{\rm BH} - \sigma_c$ relation (Equation 2.24),

$$r_{\rm inf} = 63(3-\gamma)\left(\frac{M_1}{10^9 M_\odot}\right)^{0.54}\,{\rm pc}, \tag{3.3}$$

where γ is typically in the range $1-2$, with $\gamma = 1$ approximately corresponding to 'core' galaxies and $\gamma = 2$ corresponding to galaxies with steep cusps (e.g., Faber et al. 1997). Hence, for a $10^9 M_\odot$ SMBH capturing a lighter SMBH, the binary formation radius is $r_{\rm inf} \sim 100\,{\rm pc}$.

Equation 3.3 illustrates the difficulty of directly observing pairs of gravitationally-bound AGN, which is the best possible method of proving the existence of a binary SMBH (Schnittman 2011). Rodriguez et al. (2006) discovered a pair of radio AGN with a projected separation of 7.3 pc and an estimated combined system mass $M_1 + M_2 = 1.5 \times 10^8 M_\odot$; these quantities correspond to the two SMBHs being gravitationally bound if Equation 3.3 is assumed. Deane et al. (2014) recently reported the discovery of three radio AGN in a single galaxy, with the closest pair separated by 140 pc on the sky. While this is the second-closest separation of any AGN pair after the Rodriguez et al. (2006) system, the inferred SMBH masses of $\sim 10^8 M_\odot$ imply that a bound binary has not formed.

An emerging method of identifying candidate binary SMBHs with sub-parsec separations is the search for quasars with broad emission lines that are significantly offset from the host galaxy velocity (e.g., Eracleous et al. 2012), with the working

hypothesis that only one of a pair of SMBHs is surrounded by a broad-line region. Eracleous et al. (2012) found 88 candidates with systemic velocity offsets of more than $1000\,\mathrm{km\,s^{-1}}$ within the $z < 0.7$ Sloan Digital Sky Survey catalogue, of which 14 had detectable systemic velocity variations on timescales of years. It is important, however, to recognise that only a small fraction of binary SMBHs are likely to be surrounded by a broad-line region (e.g., Yu et al. 2011), and that various other phenomena, such as a single recoiling SMBH and entrained molecular jets, can also mimic offset emission lines. Sub-parsec binary SMBH candidates may also have periodic variations in their lightcurves and precessing jets; a few promising sources appear to combine both these phenomena (Kudryavtseva et al. 2011). The century-long optical lightcurve of the quasar OJ287 has been modelled as a binary SMBH system (Valtonen et al. 2008), with a $1.8 \times 10^{10} M_\odot$ SMBH surrounded by an accretion disk that is punched through every 12 yr (as seen from the Earth) by a $1.4 \times 10^8 M_\odot$ SMBH on an eccentric ($e = 0.65$) orbit. While the orbital model, which includes post-Newtonian corrections, has been successful in predicting the optical lightcurve, little independent evidence exists for the existence of a binary SMBH in OJ287 (e.g., Agudo et al. 2012; Tanaka 2013). A number of similarly marginal claims of binary SMBH discoveries have been made; the Sudou et al. (2003) report of a particularly massive ($M_1 = 4.9 \times 10^{10} M_\odot$, $M_2 = 5 \times 10^9 M_\odot$) binary in the radio quasar 3C66B was shown to be incorrect given the lack of associated GW emission (Jenet et al. 2004).

Despite the generally inconclusive evidence for the existence of binary SMBHs available to date, the theoretical motivation for the existence of such systems, and the understanding of their behaviour, is well developed. In §3.1.1, I outline the theory of the formation of binary SMBHs in the remnants of galaxy mergers, and I provide expressions for the gravitational radiation from binary SMBHs prior to coalescence in §3.1.2. I summarise the timescales relevant to the coalescence of binary SMBHs in galaxy mergers in §3.1.3. I present the theory of isotropic, stochastic and unpolarised GW backgrounds in §3.1.4; as I shall show in forthcoming Chapters, a GW background generated by binary SMBHs is likely the most promising type of signal to search for.

3.1.1 Forming binary SMBHs in galaxy mergers

Evidence for interacting and merging galaxies is seen throughout cosmic time (Barnes & Hernquist 1992); close galaxy pairs and morphological irregularities such as tails have long been interpreted as dynamical disruptions caused by mergers (e.g., Toomre & Toomre 1972). As discussed in Chapter 2, mergers are thought to be integral to the evolutionary histories of the galaxy and SMBH populations.

The dynamics of galaxy mergers have been studied extensively over the past 50 years (Barnes & Hernquist 1992), with modern SPH n-body simulations having self-consistently followed merger evolution in a variety of scenarios (e.g., Bournaud et al. 2005; Robertson et al. 2006; Cox et al. 2008; Khan et al. 2012). Dynamical friction (Chandrasekhar 1943) acts on individual stars because of fluctuations in the gravitational fields along their orbits; the resulting deceleration is given by

$$\dot{v}_{\rm orb} = -4\pi G^2 M_{\rm star} \rho_*(< v_{\rm orb}) v_{\rm orb}^{-2} \ln(\Gamma), \tag{3.4}$$

where $M_{\rm star}$ is the mass of the star with orbital speed $v_{\rm orb}$, $\rho_*(< v_{\rm orb})$ is the density of background stars with velocities less than $v_{\rm orb}$ and $\Gamma \equiv b_{\rm max}/b_{\rm min}$ is the Coulomb logarithm, which is equivalent to the ratio of the largest to the smallest impact parameters of field stars. In galaxy mergers of reasonably high mass ratios (typically $q > 0.01$), the fluctuations of the gravitational potential at every point in space are large enough so as to rapidly mix different regions of the stellar phase space and cause relaxation over a few galactic dynamical times, which is defined as $t_{\rm dyn} = \overline{r}/\sigma_c \sim 10^8\,{\rm yr}$ (Lynden-Bell 1967). Here, $\overline{r}$ is the mean stellar radius and the latter approximation corresponds to a Milky-Way-sized galaxy. This process of violent relaxation causes the memory of the initial orbits to be lost and a maximal increase of the entropy of the system. That is, the merger of two galaxies of not too disparate masses, even if both are spirals, will result in at least a significant fraction of the stars attaining a spheroidal distribution within a few times $10^8\,{\rm yr}$. While this discussion is necessarily significantly oversimplified, numerical simulations (e.g., Robertson et al. 2006) confirm that full stellar relaxation occurs within a few Gyr.

Hence, the problem of binary SMBH formation in the mergers of massive galaxies is somewhat well-posed (Begelman et al. 1980). The dense stellar distributions surrounding each SMBH merge and violently relax on a few galactic dynamical timescales. On a significantly shorter dynamical friction timescale given by (e.g.,

Begelman et al. 1980; Yu 2002)

$$t_{\rm dyn,SMBH} = \frac{4\times10^6}{\ln N_c}\frac{\sigma_c}{200\,{\rm km\,s^{-1}}}\left(\frac{r_c}{100\,{\rm pc}}\right)^2\frac{10^8 M_\odot}{M_2}\,{\rm yr}, \tag{3.5}$$

where r_c is the radius of the newly formed stellar distribution containing N_c stars, the SMBHs sink to the bottom of the potential well and are surrounded by a stellar cusp (Bahcall & Wolf 1976). As discussed above, a bound binary forms when the orbit of the smaller SMBH is within the influence radius, $r_{\rm inf}$, of the larger SMBH; this radius is approximated by Equation 3.3. While dynamical friction continues to drive the orbital separation below $r_{\rm inf}$, this persists as long as the binding energy of the binary is the smaller than the total kinetic energy of stars on orbits comparable to the binary separation. When the condition is violated, at a binary separation of (Quinlan 1996)

$$a_h \approx \frac{GM_2}{4\sigma_c^2} \sim 27\frac{M_2}{10^9 M_\odot}\left(\frac{\sigma_c}{200\,{\rm km\,s^{-1}}}\right)^{-2}\,{\rm pc}, \tag{3.6}$$

a new mechanism is needed to facilitate orbital decay. A likely mechanism is the three-body slingshot scattering of stars on orbits which approximately intersect the binary SMBH (Saslaw et al. 1974; Frank & Rees 1976). However, as first shown by Begelman et al. (1980), the number of such stars on 'loss cone' orbits[1] in a spherical stellar background distribution is insufficient to drive orbital decay to the point where gravitational radiation can cause coalescence within a Hubble time. This apparent stalling of binary SMBHs at separations where they will apparently never coalesce is known as the final parsec problem (Milosavljević & Merritt 2003). It was, however, not clear if binary SMBHs would stall in more realistic (non-spherical) stellar distributions; as I will discuss below, the stellar distribution in fact has a large impact on the final parsec problem.

Quinlan (1996) studied the evolution of binary SMBHs within fixed spherical stellar backgrounds with Maxwellian velocity distributions by averaging over many realisations of numerical three-body 'scattering experiments'. For each random experiment, the energy and angular momentum change of the binary and the incident star were recorded, and the results were averaged over the Maxwell distribution in

[1]The loss cone is the region in the energy − angular momentum phase space occupied by stars with angular momenta such that their pericentres lie within a small factor of the binary orbit.

stellar velocity v and velocity dispersion σ_c,

$$f_{\rm Maxwell}(v,\sigma_c^2) = (2\pi\sigma_c^2)^{-3/2}\exp(-v^2/2\sigma_c^2). \tag{3.7}$$

Three dimensionless quantities were used to describe the evolution of a binary with component masses $M_1 \geq M_2$, semi-major axis a and eccentricity e:

- the binary hardening rate
$$H = \frac{\sigma_c}{G\rho_c}\frac{d}{dt}\frac{1}{a}, \tag{3.8}$$
where ρ_c is the local stellar density.

- the stellar mass ejection rate
$$J = \frac{1}{M_1+M_2}\frac{dM_{\rm ej}}{d\ln(1/a)}, \tag{3.9}$$
where $M_{\rm ej}$ is the ejected stellar mass.

- the eccentricity growth rate
$$K = \frac{de}{d\ln(1/a)}. \tag{3.10}$$

Quinlan (1996) found that H is approximately constant for $a < a_h$ (see Equation 3.6), whereas K increases approximately log-linearly with eccentricity and reducing semi-major axis and J increases log-linearly with reducing semi-major axis alone for $a < a_h$ (see also Sesana et al. 2006). Now, parameterising $J \approx A(a/a_0)^{-\alpha}$ with A, a_0 and α as free parameters, and integrating Equation 3.9, the ejected mass for a binary evolving from $a = a_h$ to a smaller semi-major axis a_2 is given by

$$\frac{M_{\rm ej}}{M_1+M_2} = \frac{A}{\alpha(a_0/a_h)^{-\alpha}}\left[\left(\frac{a_2}{a_h}\right)^{-\alpha} - 1\right]. \tag{3.11}$$

For essentially all mass ratios, $A \sim 0.2$, and for mass ratios $q = M_2/M_1 \gtrsim 1/3$, $a_0 \sim 1.7a_h$ and $\alpha \sim 0.17$ (Sesana et al. 2006). Then, for $a_2 = 0.01a_h$, $M_{\rm ej}/(M_1+M_2) \sim 1.5$. This, by definition, is greater than the mass in stars within the influence radius of the binary system! The mass of stars within the loss cone is also of the order of the reduced mass of the binary, $\mu = M_1M_2/(M_1+M_2)$ (Merritt & Milosavljević 2005). This illustrates the final parsec problem: some means of increasing the numbers of

stars that can interact with binaries, compared to the model of a fixed spherical stellar background, is required.

The final parsec problem, however, has been solved in a multitude of ways, both through purely stellar dynamical mechanisms as well as by invoking the actions of circumbinary gas (Merritt & Milosavljević 2005; Colpi 2014). From a stellar dynamics perspective, various suggestions exist for mechanisms to either replenish the loss cone or increase its share of stars. Loss cone replenishment may occur through the two-body collisional relaxation of stars ejected by a binary (Begelman et al. 1980; Yu 2002). However, Yu (2002) showed that for realistic stellar distributions at the centres of nearby galaxies, the relaxation timescale is too long to keep the loss cone full. In smaller galaxies with higher central densities, this may however prove a viable mechanism for continued binary hardening (Milosavljević & Merritt 2001). The Brownian wandering of a binary from the centre of a stellar distribution is another loss cone replenishment mechanism which is again of particular importance in smaller galaxies (Quinlan & Hernquist 1997; Merritt 2001). For major mergers of massive galaxies, however, Yu (2002) further suggested that the observed flattened, non-axisymmetric stellar distributions at the centres of galaxies that were presumably formed through mergers may offer a more populated loss cone than the spherically-symmetric stellar distributions previously considered. Merritt & Poon (2004) demonstrated the existence of sufficient centrophilic orbits in self-consistent models of triaxial galaxies to drive binary coalescence within a Hubble time. The story was essentially completed by Preto et al. (2009), who presented preliminary simulations of 'dry' galaxy mergers that resulted in triaxial stellar distributions. Preto et al. (2011) and Khan et al. (2011) together demonstrated that numerical simulations of the merger of spherical stellar distributions centred on SMBHs of various mass ratios, when placed on eccentric approach orbits, resulted in the SMBHs reaching separations where GWs could drive coalescence within a Hubble time. These results were extended by Khan et al. (2012), who showed that for mass ratios $q \lesssim 0.05$ the significantly smaller triaxiality of the merger remnants significantly reduced the binary hardening efficiency. However, despite claims by these authors of the non-dependence of their results on the number of simulation particles, their results were questioned by Vasiliev et al. (2014), who found significant number dependence in similar-resolution studies of independent triaxial stellar distributions as well as non-full loss cones. Hence, while numerical simulations of

merging galaxies *can* result in the final parsec problem being solved, further work is required to understand exactly why, and how to generalise these results to real galaxies.

Results from studies of gas dynamics in 'wet' galaxy mergers have followed a similar developmental trajectory (Colpi 2014). Using a semi-numerical approach, Armitage & Natarajan (2002) showed that a binary SMBH with a mass ratio $q \ll 1$ would coalesce in tens of Myr if embedded in a geometrically-thin gas disk with a mass of a few per cent of the total SMBH masses. Escala et al. (2004) employed SPH simulations of equal-mass-ratio binary SMBHs embedded in spherically-symmetric Bondi accretion flows, with rough correspondence to observations of ULIRGs, to show that coalescences were again possible in tens of Myr. In both cases, dynamical friction of the SMBHs against gaseous backgrounds was the orbital decay mechanism; Escala et al. (2004) found ellipsoidal gas density enhancements situated behind the SMBHs which induced gravitational drag. Numerical simulations of high-mass-ratio binaries embedded in gas disks have also resulted in efficient drag at sub-parsec separations (e.g., Escala et al. 2005; Dotti et al. 2006). However, while larger-scale simulations of merging gas-rich disk galaxies have demonstrated that parsec-scale binary formation is efficient (Mayer et al. 2007; Van Wassenhove et al. 2012; Roškar et al. 2014), no simulation has been able to follow SMBHs in merging gas-rich galaxies until they reach separations where GW-driven orbital decay can occur within a Hubble time.

3.1.2 Gravitational radiation from binary systems

Observational astrophysics with GWs promises to provide a series of shifts of paradigm and perspective in our understanding of the Universe. GWs in the General Theory of Relativity (GR; Einstein 1916) are small travelling perturbations to the spatial parts of the space-time metric that are natural consequences of Einstein's field equations. In linearised GR, the metric, $g_{\mu\nu}$, may be written as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{3.12}$$

where $\eta_{\mu\nu}$ is the Minkowski flat-space metric and $h_{\mu\nu}$, with $|h_{\mu\nu}| << 1$, is the perturbation or 'strain' tensor. In this case, Einstein's field equations may be written

as (Thorne 1980):

$$\Box h^{\alpha\beta} = -16\pi T^{\alpha\beta}, \tag{3.13}$$

where $T^{\alpha\beta}$ is the stress-energy tensor and $\Box \equiv -\partial_t^2 + \triangle$ is the d'Alembertian operator. In this thesis, I consider only the far-field regime, where the source may be treated as a point and the dynamical motions of the source are typified by $v_{\rm src} \ll c$. Then, Equation 3.13 becomes

$$\Box h^{\alpha\beta} = 0, \tag{3.14}$$

which is a homogeneous wave equation.

The solution that I shall use for the strain tensor is found using the post-Newtonian expansion of Equation 3.13 (Blanchet 2014) in $v_{\rm src}/c$. To lowest order, the wave strain three-tensor h_{ij} (we signify dropping the time-component of $h_{\mu\nu}$ by shifting to Roman indices) depends only on the distance, D, and mass quadrupole moment, Q_{ij}, of the source. Q_{ij} is defined by:

$$Q_{ij} = \int x_i x_j dm \tag{3.15}$$

where the integral runs over the source mass distribution, and the wave strain is then given by

$$h_{ij} = \frac{2G}{c^4 D}\ddot{Q_{ij}}, \tag{3.16}$$

where D is the source distance. The strain tensor h_{ij} can always be written in a transverse-traceless (TT) gauge by projecting it into a Cartesian coordinate system with the direction of propagation along the positive z-axis, and removing the trace. Then,

$$h_{ij} = \begin{pmatrix} h_+ & h_\times & 0 \\ h_\times & -h_+ & 0 \\ 0 & 0 & 0 \end{pmatrix} = h_+ \epsilon_{ij}^+ + h_\times \epsilon_{ij}^\times, \tag{3.17}$$

where I identify the linearly-independent 'plus' (+) and 'cross' (×) GW polarisation tensors

$$\epsilon_{ij}^+ = 2^{-1/2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \epsilon_{ij}^\times = 2^{-1/2} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{3.18}$$

These define the quadrupolar nature of GWs. As illustrated in Figure 3.1, polarised

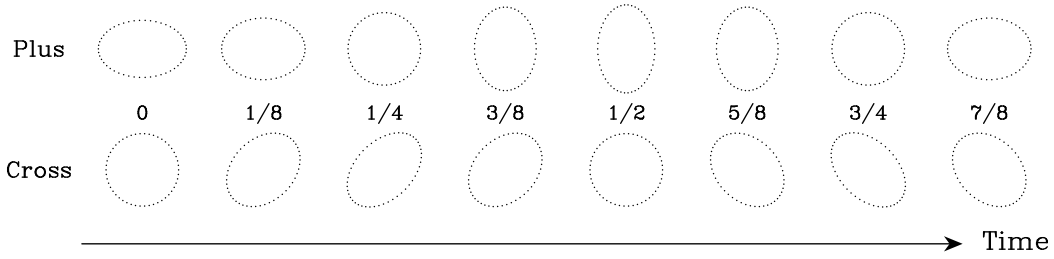

Figure 3.1: Effect on a circular ring of test masses of a GW with a strain amplitude of unity and equal polarisation amplitudes propagating out of the page. Images in the top row correspond to the + polarisation, and images in the bottom row correspond to the × polarisation. Each column corresponds to a different phase of a GW period, as labelled by the fractions.

GWs perturb a ring of test particles in free space by stretching the distribution first along one axis, and then along the perpendicular axis; the plus and cross perturbations are $\pi/4$ rotations of one another. The polarisation amplitudes, h_+ and $h_\times$, encode the time-variability and the source characteristics.

In general, GWs are searched for by directly measuring time-varying strains, which are fractional distance changes over fixed distances (Sathyaprakash & Schutz 2009). A source of GWs must have components of its motion that are neither spherically nor cylindrically symmetric, corresponding to a non-zero third time derivative of the quadrupole moment. However, the extremely small GW strain amplitudes emitted even by the most energetic GW sources in the Universe, which are binary SMBHs, make them very difficult to detect. A simple order-of-magnitude estimate of the GW strain amplitude may be attained by recognising that any component of the tensor $\ddot{Q}_{ij}$ will be of the order $(Mv^2)_{\rm non-sph}$, which is twice the non-spherical kinetic energy of the source (Sathyaprakash & Schutz 2009). Then, an upper bound on the strain amplitude of a source is

$$h \lesssim \frac{2R_S[(M(v/c)^2)_{\rm non-sph}]}{D} \approx 1.9\times10^{-19}\left[\left(\frac{v}{c}\right)^2\frac{M}{M_\odot}\right]_{\rm non-sph}\left(\frac{\rm Mpc}{D}\right). \quad (3.19)$$

This is very small! A coalescing pair of $1.4M_\odot$ neutron stars in the Virgo cluster, at a distance of 18 Mpc, will emit a maximum possible GW strain amplitude of $h \sim 1.5\times10^{-20}$. Over the diameter of the Earth, this corresponds to a distance change of less than a few hundred proton radii.

Since the Weber (1967) experiments, a multitude of GW detectors have been

built (Armstrong 2006; Hobbs et al. 2010b; Pitkin et al. 2011), but which have all invariably returned null results. Weber bars, consisting of large accurately-machined aluminium spheres or cylinders resonant at fixed kHz frequencies over small bandwidths, were the earliest detectors to be deployed. Weber (1969) claimed the detection of multiple events with $h > 10^{-16}$ that were coincident across different detectors spaced by $\sim 1500\,\mathrm{km}$; however, these claims were refuted by later experiments (Levine 2004). The largest amount of effort and funds in GW detection have been directed towards ground-based interferometer experiments, such as the currently-under-construction Advanced Laser Interferometer Gravitational-wave Observatory (Advanced LIGO, e.g., Harry & LIGO Scientific Collaboration 2010) and Advanced VIRGO (Acernese et al. 2006) facilities. These instruments will be sensitive to GWs with $h \gtrsim 2 \times 10^{-24}$ at frequencies of a few tens to a few thousands of Hz, and are likely to detect GWs within a few years of operation. Interferometer detectors use Michelson laser interferometers to measure orthogonal relative path-length changes within cavities that are a few kilometres long. The proposed Evolved Laser Interferometer Space Antenna (eLISA; Amaro-Seoane et al. 2012), with a possible launch in 2034, will create a 0.01 AU-scale interferometer using three spacecraft, which will be sensitive to GWs with $h \gtrsim 10^{-21}$ at milli-Hertz frequencies. Other space-based experiments (Armstrong 2006) use Doppler tracking measurements of interplanetary spacecraft to search for GWs in the $10^{-6} - 10^{-4}\,\mathrm{Hz}$ frequency band; the current strain sensitivity is $\sim 10^{-16}$. Finally, pulsar timing array (PTA) experiments (Hobbs et al. 2010b, also see §3.2), which form the basis for this thesis, use decade-long timing observations of millisecond radio pulsars in our Galaxy to search for GWs with frequencies in the range $10^{-9} - 10^{-7}\,\mathrm{Hz}$, with current sensitivities of $h \gtrsim 10^{-14}$.

Gravitational radiation from wide binary SMBHs

In this thesis, I am primarily concerned with GW emission from wide binary SMBH systems evaluated to first order in v_{src}/c; comprehensive calculations are presented by Peters & Mathews (1963) and Wahlquist (1987). I use these results to describe the evolution of binaries to separations $a = 3R_S$, where R_S (Equation 2.1) is the Schwarzschild radius of the more massive SMBH and $3R_S$ corresponds to the innermost stable circular orbit of a non-rotating SMBH (Hughes 2002). At these

small separations, however, the assumption of $v_{\rm src} \ll c$ breaks down. Consider the Keplerian relation between the semi-major axis, a, and the orbital frequency, $f_{\rm orb}$:

$$f_{\rm orb} = \frac{1}{2\pi}\left(\frac{G(M_1+M_2)}{a^3}\right)^{1/2}. \tag{3.20}$$

This implies a characteristic system velocity of $v_{\rm orb}/c \sim (2a/R_S)^{-1/2}$. Fortunately, the experiments that I shall discuss to detect GWs primarily focus on systems with orbital periods greater than $\sim 1\,{\rm yr}$. For a particularly massive binary with $M_1 + M_2 = 10^{10}M_\odot$, $v_{\rm orb}/c \lesssim 10^{-6}$ for $f_{\rm orb} > (1\,{\rm yr})^{-1}$. The GW waveform calculations that I use are expected to be sufficiently accurate in this regime, and the SMBHs spins, which I do not account for, are also not expected to significantly affect my results (Blanchet 2014).

For a circular binary SMBH located at a cosmological redshift z, the GW polarisation amplitudes are given by

$$h_+(t) = h_B\cos(2\pi f t + \phi_0)\frac{1+\cos^2\iota}{2} \tag{3.21}$$

$$h_\times(t) = h_B\sin(2\pi f t + \phi_0)\cos\iota, \tag{3.22}$$

where I use a coordinate system with the binary located along the z-axis and the ascending node of the orbit aligned with the x-axis, and where

$$f = \frac{f_r}{(1+z)} = \frac{f_{\rm orb}}{2(1+z)} \tag{3.23}$$

is the observer-frame GW emission frequency in terms of the rest-frame GW frequency f_r, ϕ_0 is a constant phase, ι is the binary inclination and

$$h_B = \frac{4(GM_C)^{5/3}}{c^4 D(z)}(\pi f(1+z))^{2/3} \tag{3.24}$$

is an amplitude term. In Equation 3.24, $M_C = (M_1M_2)^{3/5}(M_1+M_2)^{-1/5}$ is the chirp mass of the binary, and $D(z)$ is the comoving distance to redshift z defined as

$$D(z) = \int_0^z \frac{c}{H(z')}dz'. \tag{3.25}$$

Equations 3.21 and 3.22 imply that a 'face-on' binary will radiate GWs of both

polarisations, whereas an 'edge-on' binary will only radiate plus-polarised GWs of one quarter the summed amplitude. An edge-on binary rotated by $\pi/4$ in the x-y plane will radiate only cross-polarised GWs.

The energy in GWs radiated by a binary SMBH per unit frequency may be derived as follows (c.f. Phinney 2001). The energy flux carried by a GW is given by

$$S(t) = \frac{c^3}{16\pi G}(\dot{h}_+^2 + \dot{h}_\times^2). \tag{3.26}$$

Applying Parseval's relation to the Fourier transforms, $\tilde{h}_{+,\times}(f)$, of $h_{+,\times}(t)$,

$$\int_{-\infty}^{\infty} S(t)dt = \frac{\pi c^3}{4G}\int_0^{\infty} f^2(|\tilde{h}_+|^2(f) + |\tilde{h}_\times|^2(f)), \tag{3.27}$$

where the negative frequency parts of $|\tilde{h}_{+,\times}|^2$ have been folded onto the positive frequency parts. I have implicitly assumed that the strain amplitudes as functions of time are statistically stationary. Now, the luminosity, $L_{\rm GW}$, of a GW source is defined as

$$L_{\rm GW}(t) = 4\pi D_L^2(z)\langle S(t)\rangle_\Omega, \tag{3.28}$$

where $D_L(z) = (1+z)D(z)$ is the luminosity distance to redshift z, and the averaging is conducted over all source orientations (or equivalently all observer positions around the source). In terms of the rest-frame source luminosity as a function of time t_r, where $\frac{dt_r}{dt} = (1+z)^{-1}$,

$$\int_{-\infty}^{\infty}\langle S(t)\rangle_\Omega dt = \frac{1+z}{4\pi D_L^2(z)}\int_{-\infty}^{\infty} L_{\rm GW}(t_r)dt_r = \frac{1+z}{4\pi D_L^2(z)}\int_0^{\infty}\frac{dE_{\rm GW}}{df_r}df_r, \tag{3.29}$$

where $E_{\rm GW}$ is the total energy radiated by the source over its lifetime. Then, by comparing Equations 3.29 and 3.27,

$$\frac{dE_{\rm GW}}{df_r} = \frac{\pi^2 c^3 D_L^2}{G} f_r^2\langle|\tilde{h}_+|^2(f_r) + |\tilde{h}_\times|^2(f_r)\rangle_\Omega. \tag{3.30}$$

By inspection of Equation 3.27 and Equation 3.30, the orientation- and polarisation-averaged rms strain amplitude of a circular binary SMBH, h_s, is defined as

$$h_s^2 = \langle h_+^2(t) + h_\times^2(t)\rangle_\Omega = \frac{df_r}{dt_r}\langle|\tilde{h}_+|^2(f_r) + |\tilde{h}_\times|^2(f_r)\rangle_\Omega. \tag{3.31}$$

The term $\frac{df_r}{dt_r}$ represents the frequency-evolution of the monochromatic GW signal, which can be derived by combining Equation 3.20 with the following consideration

$$\frac{d}{dt_r}\frac{GM_1M_2}{2a} = L_{\rm GW}(t_r), \tag{3.32}$$

which results in

$$\frac{df_r}{dt_r} = \frac{96}{5}c^{-5}\pi^{8/3}f_r^{11/3}(GM_C)^{5/3}. \tag{3.33}$$

Now, by direct averaging of a form of Equations 3.21 and 3.22 which include an arbitrary rotation, ϕ, of the line of nodes, over t, ϕ, ι and ϕ_0, it is possible to obtain

$$\begin{aligned} h_s &= \left(\frac{32}{5}\right)^{1/2}\frac{(GM_C)^{5/3}}{c^4D(z)}(\pi f_r)^{2/3} \\ &\approx 7.6\times10^{-15}\left(\frac{M_C}{10^9M_\odot}\right)^{5/3}\left(\frac{100\,{\rm Mpc}}{D}\right)\left(\frac{f}{(1\,{\rm yr})^{-1}}\right)^{2/3}, \end{aligned} \tag{3.34}$$

where the latter approximation assumes $z \ll 1$. Finally, by combining Equations 3.30, 3.31 and 3.34, the energy emitted per unit frequency by a circular binary SMBH is

$$\frac{dE_{\rm GW}}{df_r} = \frac{\pi}{3G}(GM_C)^{5/3}(\pi f_r)^{-1/3}. \tag{3.35}$$

An eccentric binary system loses energy to GWs at a faster rate than a circular binary system with the same semi-major axis, and radiates GWs at multiple harmonic frequencies rather than only at $2f_{\rm orb}$ (Peters & Mathews 1963). The luminosity in GWs emitted by a binary with eccentricity e is given by

$$L_{\rm GW}(M_1, M_2, f_{\rm orb}, e) = L_{\rm GW}(M_1, M_2, f_{\rm orb})F(e), \tag{3.36}$$

where $L_{\rm GW}(M_1, M_2, f_{\rm orb})$ is defined in Equation 3.28 and

$$F(e) = \frac{1 + \frac{73e^2}{24} + \frac{37e^4}{96}}{(1-e^2)^{7/2}}. \tag{3.37}$$

The distribution of the luminosity among integer harmonics, $nf_{\rm orb}$, is defined by a function, $g(n, e)$, which I do not reproduce here for reasons of brevity, but which may be found in Equation (20) of Peters & Mathews (1963).

In the framework presented here, the time to coalescence of a circular binary SMBH evolving under the loss of energy and angular momentum to GWs alone can

be calculated by integrating Equation 3.33:

$$t_{\rm GW} = (3.2 \times 10^6\,{\rm yr}) \left(\frac{M_C}{10^9 M_\odot}\right)^{-5/3} \left(\frac{f_{\rm orb}}{10^{-9}\,{\rm Hz}}\right)^{-8/3}. \tag{3.38}$$

This timescale will be less by a factor of the order $F(e)$ for eccentric binaries.

Close to binary coalescence, numerical relativity simulations (e.g., Baker et al. 2006a) provide accurate GW templates for the emission both pre- and post-merger. Of particular interest here is the expectation of a velocity kick imparted to SMBHs newly formed from coalescing unequal-mass or spinning SMBHs (Baker et al. 2006b), which may cause ejection from haloes even at low redshifts (Gerosa & Sesana 2014). The emission of a GW 'memory burst' (Christodoulou 1991), which is a propagating step-change in the metric upon binary coalescence and which may be detectable with PTAs (e.g., Cordes & Jenet 2012), is another important effect that has been characterised through numerical simulations (Favata 2009).

3.1.3 Summary of timescales

The various stages in the formation and coalescence of binary SMBHs in galaxy mergers, and their related timescales, may be summarised as follows. I consider in particular major mergers of massive galaxies with stellar masses $M_* \gtrsim 10^{10} M_\odot$.

The galaxy merger. The initial merger is complete in a few dynamical times given by $t_{\rm dyn} \sim 10^8\,{\rm yr}$.

Sinking of the SMBHs. The SMBHs in the merging galaxies will sink to the centre of the merger remnant through stellar and/or gas dynamical friction on a timescale $t_{\rm dyn,\,SMBH} \sim 10^6\,{\rm yr}$ (Equation 3.5). Gravitationally-bound binaries are formed on this timescale.

Star- or gas-driven binary hardening. For a binary in a fixed stellar background with a continually-full loss cone, the hardening timescale is again a few Myr. This is similar to the hardening timescale in the case of a circumbinary disk (e.g., Roedig & Sesana 2012).

GW-driven binary coalescence. The GW coalescence timescale is $t_{\rm GW} \sim$ few Myr (Equation 3.38).

Hence, unless a binary SMBH formed in a major merger of massive galaxies does not undergo sufficient star- or gas-driven hardening, the SMBHs will coalesce on essentially a few galactic dynamical times.

3.1.4 The GW background from binary SMBHs

Defining a GW background

In this thesis, I am primarily interested in modelling the summed GW signals from the binary SMBH population, which are commonly assumed to together form a stochastic, isotropic and unpolarised GW background (GWBs; Maggiore 2000). GWBs are commonly represented as fractions of the critical or closure energy density of the Universe, $\rho_c c^2$, per logarithmic frequency unit:

$$\Omega_{\rm GW}(f) = \frac{8\pi G}{3H_0^2 c^2} f \frac{d\rho_{\rm GW}}{df}, \tag{3.39}$$

where ρ_{GW} is the mean energy density of the GWB in the Universe. Now, for a sky filled with GW sources, let the polarisation amplitudes corresponding to a direction vector, $\mathbf{\Omega}$, from the Earth be $\tilde{h}_{+,\times}(f, \mathbf{\Omega})$ (see Equation 3.27). The assumptions of isotropy and lack of polarisation imply that I can define

$$h_s^2(f) = 2\langle \sum_{p=+,\times} \langle |\tilde{h}_{+,\times}(f, \mathbf{\Omega})|^2 \rangle_{\phi_{\rm pol}} \rangle_{\mathbf{\Omega}}, \tag{3.40}$$

which is the mean squared GW amplitude averaged over the sky and over all polarisation position angles $\phi_{\rm pol}$. Then, I can further define a GW strain spectral density, $S_h(f)$, as

$$S_h(f) = 4\pi h_s^2(f). \tag{3.41}$$

Using this definition, $\Omega_{\rm GW}(f)$ can be expressed as

$$\Omega_{\rm GW}(f) = \frac{2\pi^2}{3H_0^2} f^3 S_h(f). \tag{3.42}$$

This can be demonstrated as follows. A GW detector can be thought of as observing the accumulation of the GW strain waveform in each polarisation over some time span $T_{\rm obs}$ at a given time t; this accumulation may be written as $h'_p(t, \mathbf{\Omega}) = h_p(t, \mathbf{\Omega}) * F_T(t)$, where $F_T(t) = 1/T$ for $-T/2 \leq t \leq T/2$ and $F_T(t) = 0$ otherwise, and

$p = +, \times$. Then, each time-derivative on the right hand side of Equation 3.26 becomes

$$\dot{h'}_p(t, \mathbf{\Omega}) = \frac{d}{dt} h_p(t, \mathbf{\Omega}) * F_T(t) = \int_{-\infty}^{\infty} 2\pi i f \tilde{h}_p(f, \mathbf{\Omega}) \mathrm{sinc}(fT) e^{2\pi i f} df. \tag{3.43}$$

where $\mathrm{sinc}(fT) = \sin(\pi fT)/(\pi fT)$. This leads to an alternate form of Equation 3.27, again using Parseval's theorem and the fact that $\tilde{h}_p(-f, \mathbf{\Omega})\mathrm{sinc}(-fT) = \tilde{h}_p(f, \mathbf{\Omega})\mathrm{sinc}(fT)$:

$$\int_{-\infty}^{\infty} S'(t, \mathbf{\Omega}) dt = \frac{c^3 \pi}{2G} \int_0^{\infty} f^2 \mathrm{sinc}^2(fT) \sum_{p=+,\times} |\tilde{h}_p(f, \mathbf{\Omega})|^2 df. \tag{3.44}$$

Now, using Equation 3.40, an estimate of the energy density of an isotropic, unpolarised GWB can be defined as

$$\hat{\rho}_{\mathrm{GW}} = \int_{-\infty}^{\infty} \frac{4\pi}{c} \delta(t) \frac{\langle S'(t, \mathbf{\Omega}) \rangle_{t, \mathbf{\Omega}}}{\mathrm{sinc}^2(fT)} dt = \int_0^{\infty} \frac{c^2 \pi^2}{G} f^2 h_s^2(f) df. \tag{3.45}$$

I assume here that clever GW observers will divide by the transfer function of their measurement setup in estimating ρ_{GW}, and I also assume that the expected value $\langle S'(t, \mathbf{\Omega}) \rangle_{t, \mathbf{\Omega}}$ is independent of time, i.e., stationary. Equating the integrand on the right hand side of Equation 3.45 to $d\hat{\rho}_{\mathrm{GW}}/df$ leads to Equation 3.42 by substitution of Equation 3.41.

Another parametrisation of a GWB is in terms of the characteristic strain spectrum, which is the rms strain amplitude per logarithmic frequency:

$$h_c(f) = \sqrt{f S_h(f)} = A_{\mathrm{yr}} \left(\frac{f}{f_{\mathrm{yr}}} \right)^{\alpha}. \tag{3.46}$$

The latter definition is a commonly-assumed power law form for the characteristic strain spectrum, with A_{yr} as the characteristic strain at a frequency of $f_{\mathrm{yr}} = (1\,\mathrm{yr})^{-1}$. An unpolarised (although not necessarily isotropic) GWB generated by a population of wide circular binary SMBHs with orbital decays driven only by GW emission will have $\alpha = -2/3$ (Phinney 2001).

The GW background from binary SMBHs

Consider a population of GW sources with comoving number density $N(z)$ at redshift z, each radiating a GW luminosity per unit rest-frame frequency, f_r, of $L(f_r)$. The specific intensity, in units of $\mathrm{W\,m^{-2}\,Hz^{-1}}$, of GWs at the Earth from sources between redshifts z and $z+dz$ is

$$dI = \frac{L(f_r)}{4\pi d_L^2}\frac{df_r}{df}N(z)\frac{d^2V_c}{d\Omega dz}dz. \tag{3.47}$$

Here, $\frac{df_r}{df} = (1+z)$, where f is the observed GW frequency, and the comoving volume element is

$$\frac{d^2V_c}{d\Omega dz} = \frac{cd_L^2}{H(z)(1+z)^2}.$$

Now, following Phinney (2001) and re-arranging, the energy density in GWs at the Earth per logarithmic frequency unit is, in any homogeneous and isotropic universe,

$$\begin{aligned}
\Omega_{\mathrm{GW}}(f)\rho_c c^2 &= \frac{4\pi}{c} f \int dI & (3.48)\\
&= \int_0^\infty \frac{fL(f_r)N(z)}{H(z)(1+z)}dz & (3.49)\\
&= \int_{t_r(\infty)}^0 \frac{f_r}{1+z}L(f_r)N(z)dt_r & (3.50)
\end{aligned}$$

where $f_r = f(1+z)$ and t_r is the proper time. The redshift z is related to t_r as

$$t_r(z) = \int_z^0 \frac{1}{H(z')(1+z')}dz'. \tag{3.51}$$

These expressions for Ω_{GW} make intuitive sense as the time-integral over the power radiated by GW sources within a comoving volume element. To extend this formulation to the binary SMBH population, further integrals are required over the distribution function of the binary SMBH population in properties such as chirp mass. In practise, the number of binary SMBHs per unit orbital separation, $N(a)$, is related to the coalescence rate of binary SMBHs, $\dot{N}$, through a continuity equation (Phinney 2001):

$$\frac{d}{da}(\dot{a}N(a)) = -\dot{N}\delta(a), \tag{3.52}$$

with solution $N(a) = -\dot{N}/\dot{a}$.

The specific case of each GW source radiating for an infinitesimal proper time corresponds to the characterisation of a GWB from circular binary SMBHs. Given that $\frac{dE_{\rm GW}}{df_r} = L(f_r)\delta(t_r)$, with $\delta(t_r)$ as the Dirac delta function, Equation 3.49 may be written as

$$\Omega_{\rm GW}(f)\rho_c c^2 = \int_0^\infty \frac{f_r}{1+z}\frac{dE_{\rm GW}}{df_r}\frac{dN(z,f_r)}{dz}dz. \tag{3.53}$$

This corresponds exactly to Equation 5 of Phinney (2001), which is a commonly-used expression in calculations of GWB energy densities, with $N(z, f_r)$ as the number of monochromatic GW sources radiating per unit comoving volume at a redshift z and a frequency f_r.

By calculating the coalescence rate of SMBHs of various masses at various redshifts, a number of predictions of $\Omega_{\rm GW}(f)\rho_c c^2$ have been made prior to this thesis (Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003a; Enoki et al. 2004; Sesana et al. 2008b). These predictions have all assumed that binary SMBHs are circular and in orbits which evolve under GW emission alone, and typically result in $\Omega_{\rm GW}(f) \propto f^{2/3}$ and $h_c(f) \propto f^{-2/3}$. The predicted range of GWB characteristic strain amplitudes, $A_{\rm yr}$, is $5\times10^{-16} - 2\times10^{-15}$, with major uncertainty caused by different model assumptions and various observational uncertainties (e.g., Sesana et al. 2008b).

3.2 Pulsar timing arrays

GWs from wide binary SMBHs are best detected with pulsar timing arrays (PTAs; Foster & Backer 1990). PTA collaborations (Manchester et al. 2013; Kramer & Champion 2013; McLaughlin 2013) measure the arrival times of radio pulses from tens of pulsars with millisecond rotation periods over five or more years, and fit these data with physically-motivated models. These models can be incredibly precise, enabling future pulse arrival times for the brightest and most stable pulsars to be predicted with $\sim$50 ns (1σ) accuracy (Manchester & IPTA 2013). This implies that no physical effects on the pulses which create delays of more than ~ 30 m exist that are not modelled; this distance is a fraction of 10^{-18} of typical kiloparsec pulsar distances. However, a GW with an rms strain amplitude $h(t)$ incident on the Earth will induce an effective Doppler shift to the measured pulsar rotation frequency, ν_p, of order $\Delta\nu_p(t)/\nu_p \sim h(t)$ (Sazhin 1978; Detweiler 1979). The induced timing

residuals, $R(t)$, which are the differences between the measured and predicted pulse arrival times, are given by

$$R(t) = \int_0^t \frac{\Delta\nu_p(t')}{\nu_p} dt' \sim \int_0^t h(t')dt'. \tag{3.54}$$

Now, if a sinusoidal $h(t) = h_0 \exp(2\pi i f t + \phi_0)$ is defined, with h_0 defined in Equation 3.34 for a binary SMBH, the rms value of the induced timing residuals is simply given by

$$\sqrt{\langle R^2(t)\rangle} = \frac{h_0}{2^{3/2}\pi f} \sim (27\,\mathrm{ns}) \left(\frac{M_C}{10^9 M_\odot}\right)^{5/3} \left(\frac{100\,\mathrm{Mpc}}{D}\right) \left(\frac{f}{(1\,\mathrm{yr})^{-1}}\right)^{-1/3}. \tag{3.55}$$

For massive, nearby binary SMBHs, this is on the order of the best pulsar timing precisions currently being achieved. However, an increase in sensitivity as well as a vital detection diagnostic is gained by comparing measured pulse arrival time data for multiple pulsars. The arrival time variations induced by GWs will be correlated between multiple pulsars (e.g., Hellings & Downs 1983), whereas any timing noise processes intrinsic to individual pulsars will not be correlated.

In this section, I begin with an outline of pulsar timing science and summarise relevant technical aspects (§3.2.1). I then quantify the pulse arrival time variations induced by GWs from binary SMBHs in §3.2.2, and conclude with a summary of important PTA results in §3.2.3.

3.2.1 Radio pulsar timing

From the very beginning of radio pulsar science (Hewish et al. 1968), studies of pulse times of arrival (ToAs) have been central. For the first pulsar to be discovered, PSR B1919+21, Hewish et al. (1968) reported an extremely stable $\sim 1.3\,\mathrm{s}$ pulse period (measured to one part in a million) as well as variations of the ToAs relative to a fixed-period model that were characteristic of Doppler shifts in the pulse rotation frequency induced by the Earth's barycentric motion. The intrinsic pulse width of 16 ms implied a source size of less than 480 km (Hewish et al. 1968); this, combined with the stability of the rotation, led Gold (1968) to infer that pulsars are rotating, magnetised neutron stars (Oppenheimer & Volkoff 1939) with zones of coherent radio emission fixed relative to the magnetic fields. The neutron star

hypothesis was spectacularly confirmed through the detection of a pulsar at the centre of the Crab nebula (Staelin & Reifenstein 1968); Pacini (1967) had previously hypothesised that the 1054 AD Crab supernova had created a rapidly spinning, magnetised neutron star, which could be the energy source powering the high-energy and optical-synchrotron nebular emission. This model stands today: neutron stars are the collapsed remnants of massive stellar supernovae within which electron degeneracy pressure is overcome by gravity to create neutron-dominated systems weighing $\sim 1.4 M_{\odot}$ with diameters of tens of kilometres.

It was soon realised that while individual pulses from pulsars are quite stochastic in intensity and shape, most pulsars have remarkably stable mean pulse profiles (e.g., Radhakrishnan et al. 1969). However, both secular and abrupt variations in pulse periods were found. In what was, to my knowledge, the first modern pulsar timing study, Cole (1969) showed that earlier observations of the spin-down of the Crab pulsar were common to four other pulsars, albeit in a weaker fashion. Cole (1969) was also able to correct the previously measured positions for these pulsars by removing annual trends from the pulsar timing residuals once the ToAs were referenced to the Solar System barycentre. Cole (1969) calculated ToAs manually on chart recordings by first measuring the time-offsets between bright pulses and a one-second system clock pulse referenced to Coordinated Universal Time (UTC) over daily few-minute-long observations, and then averaging the results to obtain mean daily ToAs. Remarkably, ToA measurement accuracies of a few milliseconds were achieved using this technique.

The analysis of pulse ToAs is the best, and in most cases the only, way to estimate fundamental parameters of pulsars. For the known population of 2302 pulsars currently listed in the Australia Telescope National Facility pulsar catalogue (Manchester et al. 2005), all have rotation period measurements and 1968 have measurements of the first period derivative, and most have timing-derived positions with 222 proper motions. Measurements of periods, P, and period derivatives, $\dot{P}$, have allowed for the classification of pulsars on the $P - \dot{P}$ diagram and the identification of a pulsar life cycle. Proper motion measurements have further inspired studies of the velocity kicks imparted to newly born pulsars in supernovae as well as the stellar velocity distribution of the Milky Way (Hobbs et al. 2005). Careful study of the characteristics of rotational glitches in pulsars (e.g., Radhakrishnan & Manchester 1969) have led to the current model for the interiors of neutron stars as hard proton-rich

crusts surrounding neutron rich superfluids, with not quite perfect crust-superfluid coupling (e.g., Warszawski & Melatos 2008). Perhaps the best-known use of pulsar timing has been in the discovery and characterisation of pulsars in binary or even multiple-object orbits. The discovery of a pulsar in a tight binary system with a second neutron star, PSR B1913+16, by Hulse & Taylor (1975), as well as the discovery of a double pulsar (Burgay et al. 2003; Lyne et al. 2004), has led to extremely stringent tests of GR in various regimes. The accurate reproduction of the orbital decay of PSR B1913+16 expected given the emission GWs resulted in the 1993 Nobel Prize in Physics being awarded to Hulse & Taylor. Numerous interesting pulsar companions have also been found, including the first extrasolar planets (Wolszczan & Frail 1992) and main-sequence stars (Johnston et al. 1992).

Today, while pulsar observing instrumentation has progressed somewhat since the early days, the essential mechanism of measuring ToAs and analysing them remains comfortingly persistent. The methods behind individual ToA measurements may be summarised as follows (e.g., Lorimer & Kramer 2012; Manchester et al. 2013).

Dedispersion. Propagation of radio pulses through the ionised interstellar medium (ISM) results in a delay with respect to infinite-frequency radiation, which is extremely well approximated by the non-relativistic sparse plasma dispersion law

$$\tau_{\rm DM} = (4.15\,{\rm ms}) \times {\rm DM} \left(\frac{\nu}{\rm GHz}\right)^{-2}, \tag{3.56}$$

where $\tau_{\rm DM}$ is the time-delay at a frequency ν in terms of the dispersion measure

$${\rm DM} = \int_0^{d_p} n_e(l) dl \,{\rm pc\,cm^{-3}}, \tag{3.57}$$

which is the line-integrated electron density, $n_e(l)$, in the pulsar direction. Typical DMs range from $2 - 1000$, depending on the pulsar location in the Galaxy (Cordes & Lazio 2002). Radio observations of pulsars typically utilise wide bandwidths of hundreds of MHz at GHz frequencies in order to increase sensitivity. Over these bandwidths, the pulses may be dispersed by many rotation periods, effectively 'smearing out' the bandwidth-integrated pulses. Two approaches may be used to mitigate this. For lower-DM pulsars observed at relatively high frequencies, it is possible to simply record time-series

data in multiple radio frequency channels and subtract the channel-dependent delays, $\tau_{\rm DM}$, in order to remove the dispersion effects. However, this technique is susceptible to pulse smearing within channels. This smearing can be fully removed by coherently filtering the Nyquist-sampled received radio-frequency electromagnetic field with the inverse of the ISM transfer function (Hankins & Rickett 1975). In both cases, the result is a time-series of measurements of the Stokes parameters of the pulsar radiation in numerous radio frequency channels, with the dispersion delays removed.

Folding and interference excision. Modern pulsar timing machines coherently fold the dedispersed signals in each radio frequency channel into single pulse profiles with thousands of phase bins. The phase bins are related to discrete time intervals by using a basic timing model, which typically includes the motion of the observatory and the pulsar spin-down. The observation start and end times are time-stamped using a local atomic clock referenced to the terrestrial time standard. Radio-frequency interference is rejected both in narrow frequency bands as well as in particularly-affected time intervals. The processes of dedispersion and folding also act as filters of certain kinds of interference. Following calibration of the radio antenna receiver characteristics, the result is a mean pulse profile in either total intensity or all Stokes parameters measured using thousands to millions of pulses and over a broad radio bandwidth, with the phase bins referenced in time to the mid-point of the observation timespan.

Template-matching. Finally, the measured pulse profile, generally in total intensity, is cross-correlated with a standard pulse profile for the specific pulsar. The standard profile may be created by summing numerous profiles, or (preferably) by fitting this sum with a series of analytic functions (Hotan et al. 2004). The standard profile is shifted to the expected pulse phase at the reference time of the observation prior to cross-correlation, and the phase offset, $\phi_R(t)$, is then measured between the standard and measured profiles. This also results an absolute pulse phase estimate, $\phi_{\rm meas}(t)$. This measurement may either be done using the usual time-domain cross-correlation, or using a Fourier-domain technique (Taylor 1992). More advanced ToA estimation techniques make use of pulse profiles in all Stokes parameters (van Straten 2006) and at multiple

radio frequencies (Pennucci et al. 2014; Liu et al. 2014).

An approximation to the 1σ uncertainty in the ToA estimation is given by (e.g., Lorimer & Kramer 2012)

$$\sigma_{\rm ToA} \sim \frac{W_{\rm eff}}{\rm SNR} \propto \frac{W_{\rm eff}}{\sqrt{\nu_p T_{\rm obs}}} \tag{3.58}$$

where $W_{\rm eff}$ is the effective pulse width in units of time, and the signal to noise ratio (SNR) of the measured pulse profile scales with the number of pulses in the measured pulse profile. Note also that the SNR scales proportionally to the intrinsic pulse flux density integrated over the observing bandwidth. Hence, fast-spinning, bright pulsars with narrow pulse widths will provide the best ToA measurements, although other factors such as ISM-driven scintillation properties and the stochasticity of the intrinsic pulse shapes also influence the ToA uncertainties (e.g., Osłowski et al. 2011; Shannon et al. 2014).

Having measured a series of ToAs, the next step is to model them. Pulsar timing models (Blandford et al. 1984; Edwards et al. 2006) include both fixed and fitted components to predict the pulse phase, $\phi_{\rm pred}(t)$, at different times. Typical models, as implemented in the commonly used software package TEMPO2 (Hobbs et al. 2006), are expressed as follows:

$$\phi_{\rm pred}(t) = \sum_{n=1}^{\rm afew} \frac{1}{n!} \frac{d^{n-1}\nu_p}{dt^{n-1}} (t - T_{\rm ref} - \Delta t)^n + \Phi_{\rm ref}. \tag{3.59}$$

Here, the $\frac{d^{n-1}\nu_p}{dt^{n-1}}$ terms are fitted to the measured pulse phases, $\phi_{\rm meas}(t)$, and typically include only the $n = 1$ (pulse period) and $n = 2$ (period derivative) quantities. The reference time, $T_{\rm ref}$, and the reference phase, $\Phi_{\rm ref}$, may be set manually. Finally, the time-offset in converting from observatory time to the pulsar emission frame, Δt, is expressed as

$$\Delta t = \Delta_C + \Delta_{E\odot} + \Delta_{R\odot} + \Delta_{S\odot} - \Delta_{\rm disp} + \Delta_B. \tag{3.60}$$

This encapsulates various time-variable physical effects, including terrestrial clock corrections (Δ_C), the Einstein time-dilation delay caused by the Earth's motion ($\Delta_{E\odot}$), the Roemer geometric propagation delay within the Solar System ($\Delta_{R\odot}$), Shapiro delays in the Solar System ($\Delta_{S\odot}$), the dispersion delay ($\Delta_{\rm disp}$) and any

delays caused by binary orbital motion of the pulsar (Δ_B). Some effects, such as the clock corrections and anything related to the Solar System ephemerides, are predetermined, whereas others, such as the contribution of pulsar position errors and proper motions to $\Delta_{R\odot}$, and possible binary motions, are fit using the measured ToAs.

The timing residuals are defined as

$$R(t) = \frac{\phi_{\rm meas}(t) - \phi_{\rm pred}(t)}{\nu_{\rm p,\,ref}}, \tag{3.61}$$

where $\nu_{\rm p,\,ref}$ is the pulsar period at a semi-arbitrarily chosen epoch. In the top three panels of Figure 3.2, I show examples of timing residuals corresponding to different timing model errors, including in the first rotation frequency derivative, the pulsar position and proper motion. In the first case, a parabolic trend is present corresponding to a rotation phase error $\phi_{\rm meas}(t) - \phi_{\rm pred}(t) \propto t^2$ (see Equation 3.59). For the position error, the annual sinusoid in the residuals corresponds to a poorly modelled Roemer delay, and the proper motion error produces a similar sinusoid modulated by the fact that the position error steadily increases with time relative to a position epoch.

Fits of the timing model are carried out by maximising the model likelihood, which is equivalent to the probability of obtaining the residuals, $R(t)$, given a particular set of model parameters. In the majority of pulsar timing applications, a weighted linear least-squares fitting algorithm is employed (Hobbs et al. 2006), with the weights determined by the formal ToA uncertainties. However, the breakdown of the implicit assumption of uncorrelated timing residuals that is observed for some pulsars, which exhibit red timing noise (Hobbs et al. 2010a; Shannon et al. 2013), has necessitated the use of generalised least-squares fitting techniques using self-consistent estimates of the covariance matrix of the residuals (Coles et al. 2011). Bayesian timing model parameter estimation is also beginning to be used (Lentati et al. 2014).

The discovery of PSR B1937+21 (Backer et al. 1982) revealed a new class of pulsar with millisecond spin periods ($P \lesssim 20\,{\rm ms}$), low spin-down rates ($\dot{P} \lesssim 10^{-17}\,{\rm s\,s^{-1}}$) and comparatively stable long-term rotation properties. The extremely weak surface magnetic fields implied by their low spin-down rates suggest that these systems are older than the bulk of the pulsar population, which have slower rotations but higher

spin-down rates. The likely scenario for the origin of the millisecond spin periods is accretion-driven spin-up, or recycling, over $\sim 10^7$ yr from an evolved stellar companion overflowing its Roche lobe (Bhattacharya & van den Heuvel 1991). Today, 279 millisecond pulsars are known, with 155 systems that are not associated with globular clusters of which 106 are in binary systems (Manchester et al. 2005). The large numbers of millisecond pulsars in globular clusters, which are old stellar systems, confirms that they represent old pulsars, and the large fraction of binary millisecond pulsars is evidence for the recycling formation model. The recent discovery of a 'missing link' millisecond pulsar switching between states of rotation-powered radio pulsations and accretion-powered X-ray pulsations (Papitto et al. 2013) on timescales of weeks also demonstrated the existence of systems in the process of being spun up through accretion.

Millisecond pulsars are the ultimate celestial clocks because of their high rotation frequencies (see Equation 3.58) and significantly lower levels of intrinsic red timing noise compared to the rest of the pulsar population (Shannon & Cordes 2010). Ironically, the first millisecond pulsar to be discovered, which has the fastest (published) rotation for pulsars outside globular clusters, also exhibits the largest amount of red timing noise among the millisecond pulsar population. Generally, the best timing precisions are attained for millisecond pulsars that are not in globular clusters or particularly complicated binary systems because of the difficulties in creating appropriate timing models. Currently, all PTA programs together observe approximately 50 millisecond pulsars approximately once every $3 - 4$ weeks (Manchester & IPTA 2013).

3.2.2 Pulsar ToA variations induced by GWs

By considering the effects of perturbations to flat space-time caused by GWs on photon propagation paths, various authors (e.g., Sazhin 1978; Detweiler 1979; Hobbs et al. 2009; van Haasteren & Levin 2010) have derived the effects of GWs of arbitrary forms on measured pulsar rotation frequencies. These calculations are generally based on the work of Estabrook & Wahlquist (1975) on the response of Doppler spacecraft tracking measurements to GWs. Fundamentally, the GW-

induced shifts to a pulsar rotation frequency may be written as

$$\frac{\Delta\nu_p(t)}{\nu_p} = \left[\frac{\Delta\nu_p(t)}{\nu_p}\right]_{\rm Earth} - \left[\frac{\Delta\nu_p(t)}{\nu_p}\right]_{\rm Pulsar}. \tag{3.62}$$

That is, the contributions from the GW perturbations at the Earth and at the pulsar can be separated into Earth and pulsar terms, which are then differenced. For a GW defined by the tensor waveform $A_{ij}^{\rm Earth}(t)$ at the Earth, and a GW defined by $A_{ij}^{\rm Pulsar}(t - c^{-1}D_p(1-\cos\theta_p))$ at the pulsar with θ_p as the angle between the pulsar and GW source directions and D_p as the pulsar distance, the observed rotation frequency shifts are given by

$$\left[\frac{\Delta\nu_p(t)}{\nu_p}\right]_{\rm Earth} = \frac{1}{2}\hat{k}_p^i\hat{k}_p^j A_{ij}^{\rm Earth}(t)(1-\cos\theta_p)^{-1} \tag{3.63}$$

$$\left[\frac{\Delta\nu_p(t)}{\nu_p}\right]_{\rm Pulsar} = \frac{1}{2}\hat{k}_p^i\hat{k}_p^j A_{ij}^{\rm Pulsar}(t - c^{-1}D_p(1-\cos\theta_p))(1-\cos\theta_p)^{-1}, \tag{3.64}$$

where $\hat{k}_p^i$ is the unit direction vector to the pulsar from the Earth. The induced ToA variations are then derived by using Equation 3.54. For wide binary SMBHs, the rate of GW frequency evolution is slow enough to imply that, for a typical PTA observation time of 10 yr, the emitted GW frequency will evolve by less than $(10\,{\rm yr})^{-1}$, which implies that they are effectively non-evolving over the observation spans. However, significant evolution may be present over the pulsar-term time retardation, $c^{-1}D_p(1-\cos\theta_p))$, which may be $\gtrsim 10^4$ yr.

The ToA variations induced by a stationary, stochastic GWB may be characterised by their power spectral density, which is related to the characteristic strain spectrum as (e.g., Jenet et al. 2006)

$$S_g(f) = \frac{1}{12\pi^2} f^{-3} h_c^2(f). \tag{3.65}$$

These ToA variations are correlated between different pulsars, because the same metric perturbations at the Earth affect all observations. However, the degree of correlation, as measured by the normalised cross-correlation statistic between ToA variations for pulsars p and q, ρ_{pq}, will depend on the angle, θ_{pq}, between the Earth-pulsar vectors for pulsars p and q as (Hellings & Downs 1983):

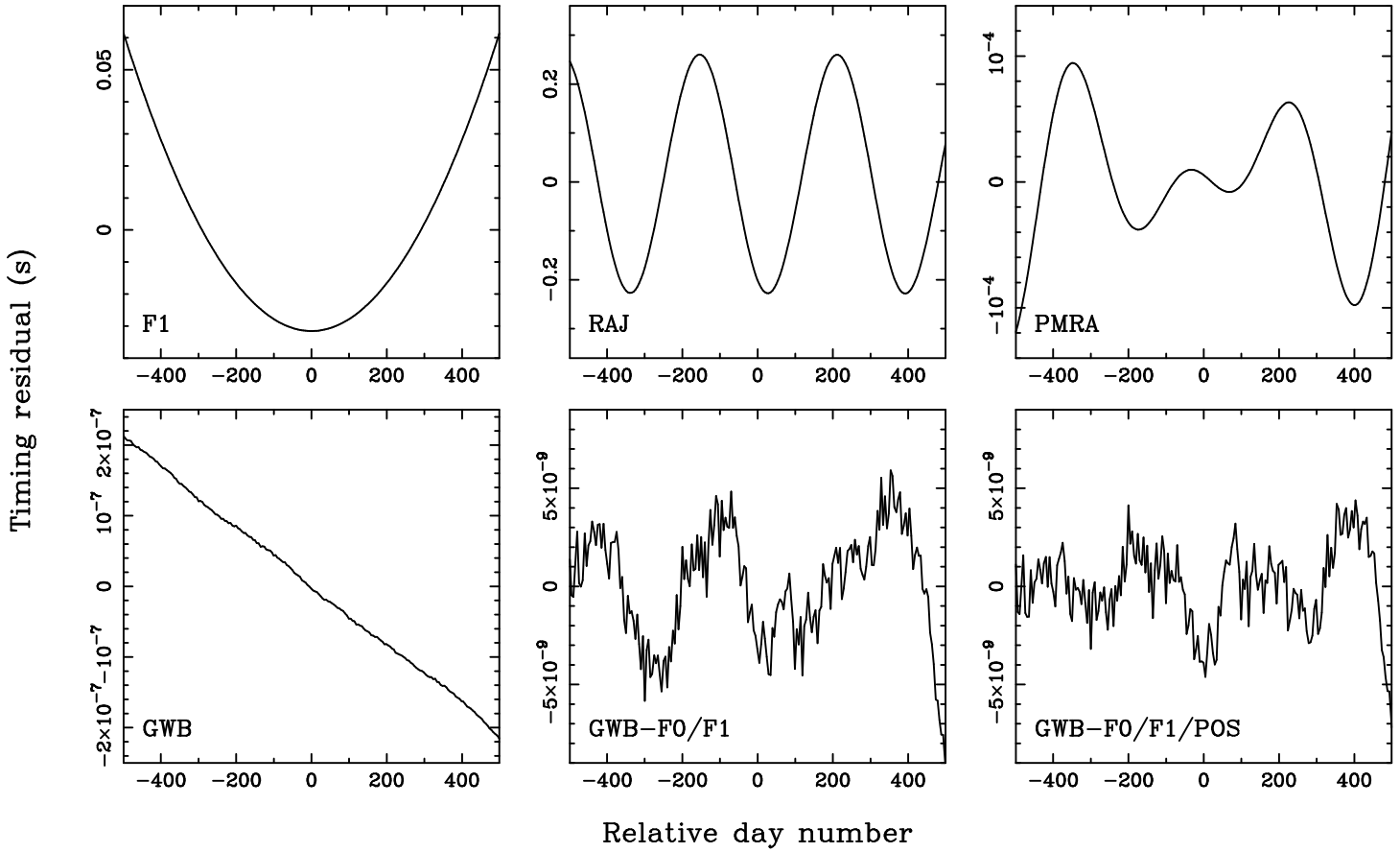

Figure 3.2: Examples of simulated pulsar timing residuals corresponding to different physical effects not included in the timing model. The simulated pulsar period is 1 s, and 1 ns rms white timing noise is present in all cases. *Panel F1:* error of $+10^{-16}\,\mathrm{s\,s^{-1}}$ in the first rotation frequency derivative relative to the correct value of $+10^{-13}\,\mathrm{s\,s^{-1}}$. *Panel RAJ:* error of $-10''$ in the pulsar right ascension. *Panel PMRA:* Error of $-40\,\mathrm{mas\,yr}-1$ in the proper motion in the right ascension coordinate. *Panel GWB:* Instead of timing model errors, the effects of a simulated stochastic GWB with $A = 10^{-15}$ and $\alpha = -2/3$ in Equation 3.46 are shown. *Panel GWB-F0/F1:* for the simulated data displayed in Panel GWB, the effects of fitting for the pulsar rotation frequency and frequency derivative are shown. *Panel GWB-F0/F1/RAJ/DECJ:* Similar to the previous panel, but with an additional fit for the pulsar position.

$$\rho_{pq} = \frac{3}{2}\beta \ln \beta - \frac{\beta}{4} + \frac{1}{2} + \frac{\delta_{pq}}{2}, \tag{3.66}$$

where $\beta = (1/2)(1 - \cos\theta_{pq})$, and the Kronecker delta, δ_{pq}, is unity if $p = q$ and zero otherwise. This correlation, known as the Hellings & Downs curve, will hold on expectation as long as the GWB is truly stochastic, with unconstrained prior knowledge of GW source orientations, polarisations and positions. The detection of this correlation is the goal of all PTA searches for a GWB.

Examples of timing residuals corresponding to the presence of a GWB are shown in the bottom three panels of Figure 3.2. By construction, the only sources of non-zero residuals in the examples are white Gaussian homoscedastic timing noise with a standard deviation of 1 ns, and a random realisation of a stochastic, isotropic, unpolarised GWB (Hobbs et al. 2009) with $A_{\rm yr} = 10^{-15}$ and $\alpha = -2/3$ (see Equation 3.46). In the bottom left panel, I show the residuals with no pulsar parameter fitting carried out: the linear trend is typical of a red noise process extending to much lower frequencies than then inverse data span. Fitting for the pulsar rotation frequency and its derivative (bottom middle panel) result in much of the red noise being absorbed; as the presence of this noise was not accounted for in the fit, the confidence intervals on the measured frequency and frequency derivative will be underestimated (Coles et al. 2011). Finally, the bottom right panel shows the results of a fit for the pulsar position as well. Figure 3.2 demonstrates some of the difficulties in studying GWs with PTAs given the necessity to additionally fit for pulsar parameters.

Throughout this thesis, I assume that only two polarisations of GWs exist (the + and × polarisations), as predicted by GR. However, recent work has shown that PTAs are extraordinarily sensitive to some polarisation states of GWs predicted by alternative theories of gravity (Lee et al. 2008; Chamberlin & Siemens 2012). A general metric theory of gravity allows for up to six independent polarisation states, corresponding to the six degrees of freedom of the Riemann curvature tensor. Experiments provide strong evidence that all theories of gravity must specify a space-time metric (Will 2014), hence constraining the number of GW polarisations to be searched for. Lee et al. (2008) and Chamberlin & Siemens (2012) both considered the detectability of a stochastic GWB radiated in each GW polarisation beyond those predicted by GR, by deriving analogues of the Hellings & Downs curve. They

found that pulsar pairs with small angular separations are up to 10^4 times more sensitive to non-GR polarisations than to GR polarisations. PTAs hence have the potential to perform stringent tests of GR by constraining the shape of a measured Hellings & Downs curve.

3.2.3 Historical overview of pulsar timing arrays

Early applications of pulsar timing towards GW science were focused on showing that the energy density in gravitational radiation was not sufficient to close the Universe (i.e., $\rho_{\rm GW} < \rho_c$). Scale-invariance of the scalar potential fluctuations during the inflationary era is predicted to give rise to a flat-spectrum $\Omega_{\rm GW}(f)$, which, from Equations 3.42, 3.46 and 3.65, implied $P_{\rm GW}(f) \propto f^{-5}$ corresponding to a steeply red noise process. Using published timing residual rms values, Detweiler (1979) was the first to demonstrate that $\Omega_{\rm GW} h_{100}^2 < 1$ with 95% confidence, implying that inflationary-era GWs alone could not provide the critical density of the Universe. This work was extended by Romani & Taylor (1983), who used a power-spectral analysis of the residuals from long-term timing measurements of PSR B1237+25 from the Jet Propulsion Laboratory (JPL) to constrain $\Omega_{\rm GW} h_{100}^2 < 5 \times 10^{-4}$. By cross-correlating timing residuals for a set of four pulsars observed by the JPL group, Hellings & Downs (1983) obtained a slightly better limit of $\Omega_{\rm GW} h_{100}^2 < 1.4 \times 10^{-4}$ by constraining the possible presence of the Hellings & Downs curve. Following the discovery of two millisecond pulsars, PSRs B1937+21 and B1855+09, Stinebring et al. (1990) were able to use their significantly enhanced long-term timing precisions to show that $\Omega_{\rm GW} h_{100}^2 < 4 \times 10^{-7}$, effectively ruling out any significant contribution of a flat-spectrum GWB to the energy density of the Universe, and beginning to constrain models of cosmic strings as the origins of matter-density fluctuations.

The first dedicated PTA program was established by Foster & Backer (1990) at the National Radio Astronomy Observatory, and included observations of three millisecond pulsars. Foster & Backer (1990) described a search for the Hellings & Downs correlations, as well as for the effects of incorrect terrestrial time standards or planetary ephemerides, which would manifest as monopole or dipole correlations as function of the pulsar angular separations. A detailed timing analysis of longer-term data from the Arecibo Observatory on these three pulsars by Kaspi et al. (1994)

placed even more stringent GWB limits, and also provided a publicly-available data set that is proving invaluable to current PTA efforts (e.g., Shannon et al. 2013). Further analysis on a similar, more extended data set by Lommen (2002) limited $\Omega_{\rm GW}h^2_{100} < 2\times10^{-9}$. Finally, Jenet et al. (2004) were able to discount the suggestion by Sudou et al. (2003) of a particularly massive binary SMBH in the radio quasar 3C66B, by showing that the GW emission expected from this object would have been observable in existing pulsar timing data sets.

Since these "early" days, the PTA field has exploded in size and scope. The establishment of the Parkes Pulsar Timing Array (PPTA; Manchester et al. 2013) in 2004 with the Parkes Telescope in Australia marked the first concerted effort to *detect* a predicted GW signal with PTAs; this signal was the GWB generated by binary SMBHs (Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003a). Soon after, the European Pulsar Timing Array (EPTA; Kramer & Champion 2013) and the North American Nanohertz Observatory for Gravitational Waves (NANOGrav McLaughlin 2013) followed suit, utilising telescopes in Europe and North America respectively. Today, these three collaborations all share data as part of the International Pulsar Timing Array (IPTA; Hobbs et al. 2010b). Searches have been conducted for GWBs with various spectral indices, individual continuous-wave sources of GWs and GW memory bursts. As shall be discussed in Chapter 4, the best current limit on the characteristic strain amplitude of the GWB at a frequency of $f_{\rm yr}$, is $A_{\rm yr} < 2.4 \times 10^{-15}$ with 95% confidence (Shannon et al. 2013). This is quite close to the range of predicted GWB amplitudes prior to this thesis ($5 \times 10^{-16} - 2 \times 10^{-15}$). In Chapter 8, I shall further demonstrate that the Shannon et al. (2013) GWB constraint is inconsistent with viable models for the binary SMBH population.

Chapter 4

The best PTA constraints on GWs from binary SMBHs

I review the best existing pulsar timing array (PTA) constraints on gravitational-wave (GW) signals from binary supermassive black holes (SMBHs), which at the time of writing have all been derived by the Parkes PTA (PPTA) collaboration. I describe an upper limit on the strength of the stochastic, isotropic GW background (Shannon et al. 2013), as well as all-sky upper limits on the amplitudes of GWs from individual binary SMBH systems. For the latter class of signal, I consider in particular continuous-wave GW emission from wide binary SMBHs (Zhu et al. 2014), and GW bursts with 'memory' from SMBH-SMBH coalescence events (Wang et al. 2015). While I was an author on each of the three aforementioned publications, I was not a major contributor to the results from these publications that I describe in this Chapter. I also review current expectations for the sensitivities of future PTA data sets to GWs from binary SMBHs.

The most sensitive PTA data set to all GW signals from binary SMBHs was recently described by Manchester et al. (2013). The data include the first release of dedicated PPTA observations ("data release 1", or DR1), combined with previous Parkes observations presented by Verbiest et al. (2009) (DR1E).

All DR1/DR1E observations were made with the CSIRO Parkes radio telescope

in rural New South Wales, Australia. The Parkes telescope is a 64 m dish equipped with a selection of prime focus receivers. The DR1 observations, which commenced on 2004 May 20, were primarily made in frequency bands around 700 MHz, 1.4 GHz and 3.1 GHz. Twenty pulsars were observed; their sky positions are shown in Figure 4.1 overlayed upon a 408 MHz map of the sky (Haslam et al. 1982), which represents the sky-variable component of the receiver system temperature. Most pulsars were observed for 1 hr in each frequency band, approximately every three weeks. Data in multiple frequency bands enable the measured ToAs to be corrected for $\sim \mu$s-scale secular variations in pulsar DMs; such corrections were applied for all DR1 data (Keith et al. 2013). Also, while most pulsars have negative power-law radio spectra, pulse widths also typically reduce with increasing frequency. This means that if sufficiently high-fidelity pulse profiles can be recorded at high frequencies, better timing precision may be obtained compared to lower-frequency data. This was the case for the three best-timed PPTA pulsars (Manchester et al. 2013). Timing residuals from the DR1 data are shown in Figure 4.2; the best-fit timing model parameters were derived using generalised least-squares fits as described in Chapter 3 (Coles et al. 2011). The previous Parkes observations presented by Verbiest et al. (2009), which were combined with the DR1 data set to form DR1E, were only made around 1.4 GHz over a smaller radio bandwidth. However, their addition was important in increasing the PPTA sensitivity to a GW background from binary SMBHs, which is expected to induce ToA variations with corresponding to a red-noise process.

Three analyses of the PPTA DR1/DR1E data have been conducted to search for and constrain the existence of GWs. Shannon et al. (2013), hereafter S13 in this Chapter, found an upper limit on the energy density, $\Omega_{\rm GW}$, of a stochastic, isotropic GW background (GWB) from binary SMBHs. This limit was six times more stringent than previous results from the EPTA (van Haasteren et al. 2011) and NANOGrav (Demorest et al. 2013). It is important to note that S13 did not report a *search* for the GWB.

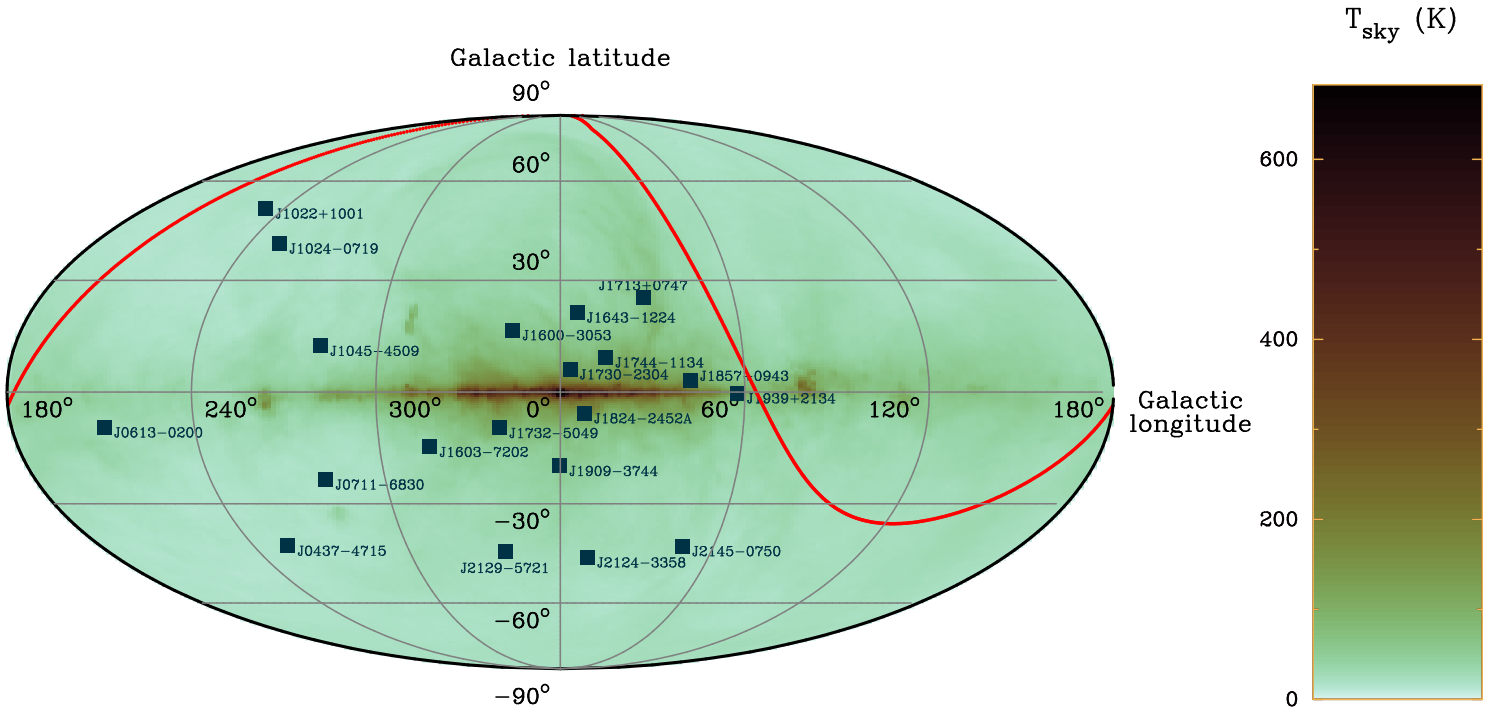

Figure 4.1: Sky positions in Galactic coordinates of the 20 pulsars observed by the PPTA, shown as blue squares. The background image is the 408 MHz sky temperature map produced by Haslam et al. (1982), with striping and point sources removed. The dominant source of radio emission in this map is diffuse synchrotron emission associated with the Milky Way. The temperature scale is given by the colour wedge to the right of the image. Also shown, as a red line, is the declination limit of the Parkes telescope; sources to the North of the red line cannot be observed from Parkes.

Zhu et al. (2014), hereafter Z14, searched for continuous-wave (CW) signals from circular, non-evolving binary SMBHs, and Wang et al. (2015), hereafter W15, searched for GW memory bursts which are expected from SMBH-SMBH coalescences. Neither search resulted in any detections, and Z14 and W15 present constraints on the all-sky occurrence of wide binary SMBHs and the all-sky rate of memory burst events respectively. Z14 improved upon previous NANOGrav results (Arzoumanian et al. 2014) by a factor of two, while the W15 study was the first search for GW memory burst events. In the following three sections, I summarise in turn the methods and results of S13, Z14 and W15.

4.1 A constraint on the GWB from binary SMBHs

S13 devised a new method of deriving an upper limit on the strength of the GWB from binary SMBHs using PTA data, which built on earlier work by Jenet et al. (2006) and Verbiest (2008). The mean characteristic strain spectrum of the GWB was assumed to take the form in Equation 3.46:

$$h_c(f) = A_{\rm yr} \left(\frac{f}{f_{\rm yr}} \right)^{-2/3} . \tag{4.1}$$

As discussed below Equation 3.46, this spectrum corresponds to the assumption that all binary SMBHs radiating GWs in the PTA frequency band, defined below for the PPTA, are in circular orbits which evolve only through losses of energy and angular momentum to GWs. S13 hence presented upper limits on $A_{\rm yr}$. The S13 study forms the basis of my comparison in this thesis between models for the binary SMBH population and PTA data.

S13 derived an optimal estimator for $A_{\rm yr}$ based on estimates of the power-spectral densities (PSD) of pulsar timing residuals. Let $\hat{P}_j(f_i)$ be the residual PSD estimates at a frequency f_i indexed by an integer i, for a pulsar indexed by an integer j. Further, let the shape of the GWB-induced residual PSD be $g_j(f_i)$; this shape is *not* a power law, as may be expected from Equation 3.65, because the effects of fitting pulsar parameters mean that the statistics of the GWB-induced residuals are not the same as the statistics of the GWB induced ToA variations (see Figure 3.2 and

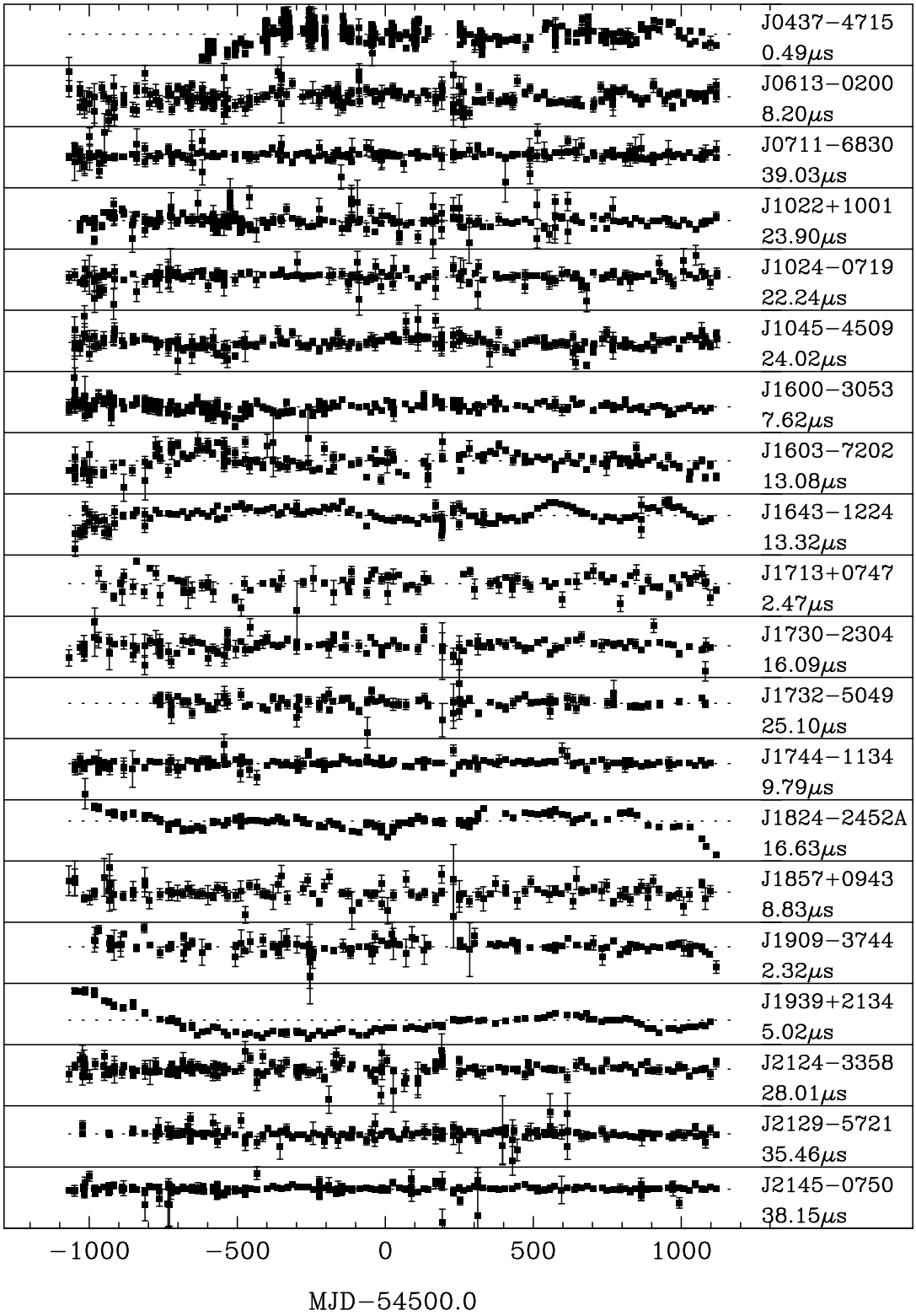

Figure 4.2: Timing residuals from the DR1 data for the 20 PPTA pulsars; the pulsar names are listed to the right of image. The vertical extent of each plot is listed in μs under each pulsar name.

related discussion). Then, a positively biased estimator of $A_{\rm yr}$ is

$$\hat{A}^2{}_{ij} = \frac{\hat{P}_j(f_i)}{g_j(f_i)}. \tag{4.2}$$

The positive bias results from the positive-definite contributions to the PSD estimates of any noise processes contributing to the residuals. An overall estimator of $A_{\rm yr}$ may then be derived from the individual $\hat{A}^2{}_{ij}$ estimates by defining weights, K_{ij}, such that the statistic

$$\hat{A}^2 = \frac{\sum_{ij} K_{ij} \hat{A}^2{}_{ij}}{\sum_{ij} K_{ij}} \tag{4.3}$$

is optimal. These optimal weights may be derived from a simple application of the Wiener filter:

$$K_{ij} = \left[\frac{g_j(f_i)}{M_j(f_i)} \right]^2, \tag{4.4}$$

where $M_j(f_i) = G_j(f_i) + W_j + R_j(f_i)$ is a smoothed model for the residual PSD containing a GWB contribution, $G_j(f_i) = A_M^2 g_j(f_i)$, a white noise contribution, W_j, and a red noise contribution, $R_j(f_i)$, described by a power law. These three contributions were estimated by S13 using a non-linear joint fit to the PSD estimates for all pulsars to model the common GWB component as well as the independent white noise components. Pulsar-specific red noise contributions were added if sufficient evidence, as determined by a likelihood ratio test, was present. Uncertainty in the determination of the noise models, $M_j(f_i)$, only causes the estimator $\hat{A}^2$ to be sub-optimal, rather than biased. The estimator is optimal only in the case of every PSD estimate, $\hat{P}_j(f_i)$, being independent and χ^2-distributed with two degrees of freedom.

The residual PSD estimates were calculated using the generalised least-squares technique of Coles et al. (2011). This technique uses an iterative process to estimate the PSD of the residuals, which is then used as a noise model to re-estimate the PSD until self-consistency is achieved. Nearly-independent PSD estimates, $\hat{P}_j(f_i)$, at harmonics of the fundamental inverse data span frequency can hence be produced. S13 further assumed that all noise processes in the residuals were Gaussian random processes, implying that the PSD estimates are χ^2-distributed. Data for only six pulsars from DR1E were used by S13 (with some modifications), as the addition of further data sets did not significantly affect the results. The modelled level of

the common GWB component was $A_M = 1.2 \times 10^{-15}$; the shapes, $g_j(f_i)$, of the GWB-induced residual PSDs were determined through simulations for each pulsar. The observed value of the detection statistic, $\hat{A}$, was $\hat{A} = 1.6 \times 10^{-15}$. The residual PSD estimates, P_j, and their models, M_j, for the six pulsar data sets used by S13 are shown in Figure 4.3.

While the values of the estimates of $A_{\rm yr}$ from the data are interesting, they are not representative of the outcome of a GWB detection process. This is because any red noise processes in the six pulsars studied by S13 with the spectral shapes approximated by $A_{\rm yr}^2 g_j(f_i)$ could correspond to the values of the $A_{\rm yr}$ estimates. An inspection of Figure 4.3 shows that the estimate of $A_{\rm yr}$ from the PSDs must be uncertain. Generally, it is accepted that any statistic used to demonstrate a detection of the GWB must search for the Hellings & Downs (1983) correlations between pulsar data sets (e.g., Jenet et al. 2005; van Haasteren et al. 2011; Yardley et al. 2011; Demorest et al. 2013), which are expected to be unique to a stochastic, isotropic GWB.

S13 placed a frequentist upper limit on $A_{\rm yr}$ by comparing the measured value of the statistic $\hat{A} = 1.6 \times 10^{-15}$ with distributions of this statistic for simulated data sets corresponding to specific values of $A_{\rm yr}$. The simulated data sets were produced with cadences and homoscedastic white ToA variation levels that were identical to the observations, along with GWB-induced ToA variations simulated using the GWbkgrd plugin (Hobbs et al. 2009) to the TEMPO2 pulsar timing package (Hobbs et al. 2006). I revisit the assumptions for the GWB implicit in this technique in Chapter 5. No pulsar-specific red noise was added, although such red noise was included in the model for PSR J1713+0747 (see Figure 4.3); this only resulted in a conservative bias for the upper limit on $A_{\rm yr}$. For each simulation, $\hat{A}^2$ was measured using the same weights K_{ij} as for the real data, and many simulations were conducted to empirically estimate the distribution of $\hat{A}^2$ for a given $A_{\rm yr}$. The value of $A_{\rm yr}$ was adjusted until a value, A_{95}, was found such that 95% of the simulated $\hat{A}^2$ statistics lay above the measured statistic. The resulting value of $A_{95} = 2.4 \times 10^{-15}$ was taken to be the conservative upper bound, with 95% confidence, on the GWB amplitude $A_{\rm yr}$. It is apparent from Figure 4.3 that an arbitrary simulation of the residual PSDs for all the pulsars lies above the measured PSDs at low frequencies, where the GWB-induced residual PSD dominates the white noise.

The S13 upper limit on $A_{\rm yr}$ was significantly lower than previous upper limits

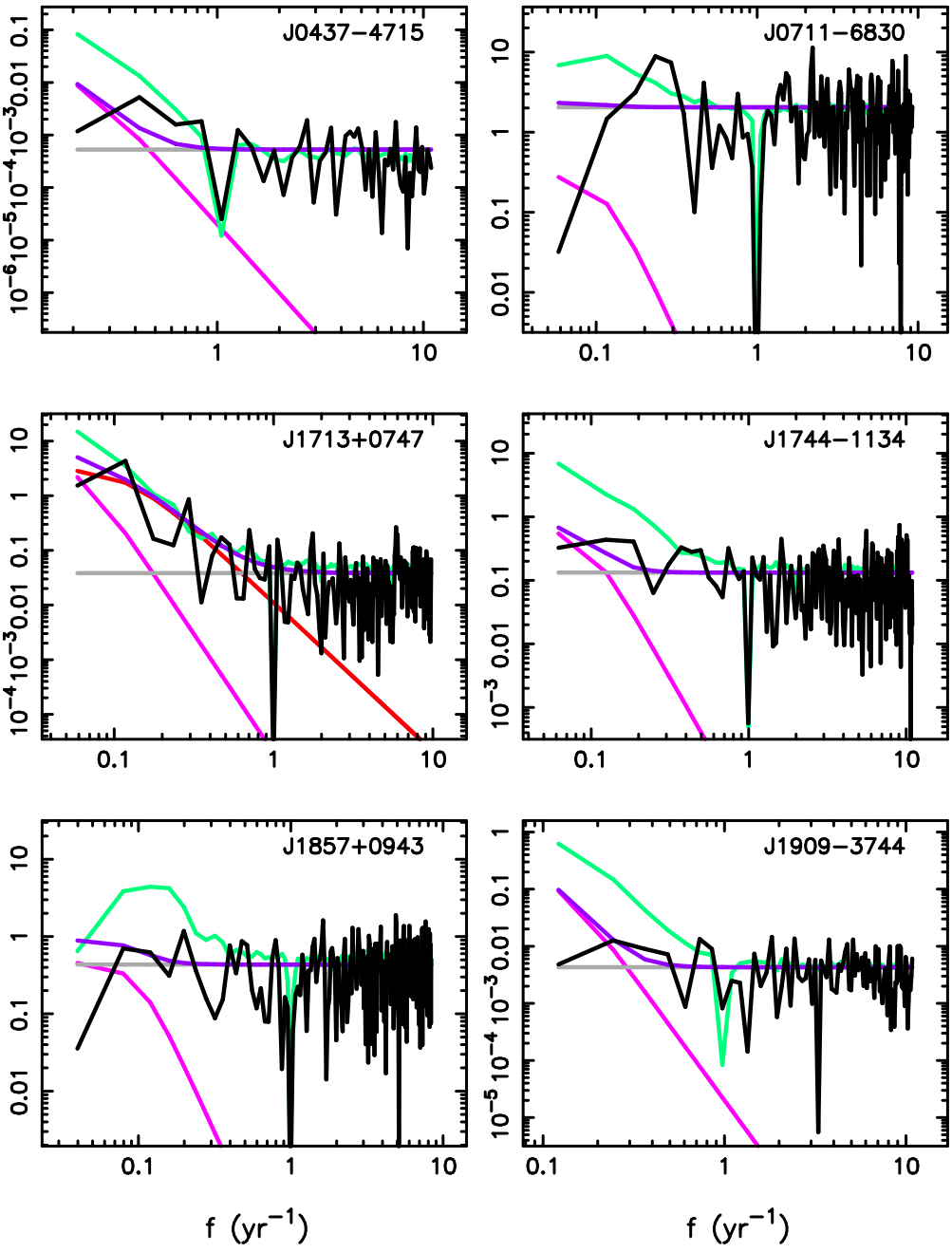

Figure 4.3: PSD estimates of the timing residuals, and models for these spectra, for the PPTA pulsars used by S13 to constrain the GWB. The PSD estimates, P_j, for the pulsars are shown as black lines, along with the models, M_j, shown as purple lines. The models contain a white component, W_j (grey lines), a common GWB component, G_j (pink lines), and, for PSR J1713+0747, an additional red noise component, R_j (red line). The green curves show what the residual PSDs would look like (on average) in the presence of a GWB with $A_{\rm yr} = 2.4 \times 10^{-15}$. The names of the pulsars are given in the top-right corners of each panel. This figure is reproduced from Shannon et al. (2013).

from the EPTA and NANOGrav of $\sim 6.5 \times 10^{-15}$. However, a better comparison between the data sets is in upper limits on $\Omega_{\rm GW}$, which are proportional to $A_{\rm yr}^2$ (see Equation 3.42). Following Demorest et al. (2013), S13 found that upper limits on $A_{\rm yr}$ for various values of α in Equation 3.46 were well fit by the curve

$$A_{\rm yr}(\alpha) = h_c(f_{\rm PPTA}) \left(\frac{f}{f_{\rm yr}} \right)^{-\alpha}, \tag{4.5}$$

where $f_{\rm PPTA} = 2.8 \times 10^{-9}\,{\rm Hz}$ is a constant (i.e., independent of α) which hence corresponds to the characteristic frequency of the PPTA constraint on the GWB. Only values of $\alpha < 3/2$ were trialled, as these values correspond to GWB-induced ToA variations with red power spectra (see Equation 3.65). The data set used by S13 was hence primarily sensitive to the GWB from binary SMBHs at a frequency of $2.8 \times 10^{-9}\,{\rm Hz}$; however, the range of GW frequencies probed by the data set is greater, as seen in Figure 4.3. Given the value of $f_{\rm PPTA}$, the PPTA constraint on $\Omega_{\rm GW}$ could be related to $A_{\rm yr}$ for $\alpha = -2/3$ as (see Equation 3.42)

$$\Omega_{\rm GW}(f_{\rm PPTA}) = \frac{2\pi^2}{3H_0^2} f_{\rm PPTA}^{2/3} f_{\rm yr}^{4/3} A_{\rm yr}^2 = 1.3 \times 10^{-9} \frac{H_0}{73\,{\rm km\,s^{-1}\,Mpc^{-1}}}. \tag{4.6}$$

The GWB amplitude that would be detectable with the PPTA DR1E data set, for example with 95% confidence, is not known. Similar calculations have also not been performed for either of the currently published EPTA and NANOGrav data sets. Instead, the best available calculation of the sensitivities of PTAs to the GWB from binary SMBHs is the analytic study of Siemens et al. (2013). These authors calculated the signal to noise ratio (S/N) of an optimal detection statistic for the GWB amplitude based solely on measuring the Hellings & Downs (1983) correlations (Anholm et al. 2009) as a function of various variables. These variables included the number of pulsars being observed, the total observation time span and cadence, and the white and red noise levels for each pulsar. The effects of stochasticity in the GWB itself contributing to the variance of the detection statistic, in particular when the GWB dominates the white noise at low frequencies, were taken into account in the S/N calculation. The characteristic strain spectrum of the GWB was assumed to be of the form in Equation 4.1. The effects of fitting timing models were crudely approximated by introducing a low-frequency cut-off to the power spectrum of the ToA variations at the inverse data span. The crudeness of this

approach can be visualised in Figure 4.3 by inspection of the estimated functions $g_j(f_i)$, which quantify the shape of the PSD of the GWB-induced timing residuals after fitting a timing model; significant differences from power law models at frequencies greater than the inverse data spans are apparent in some cases (e.g., PSRs J0711−6830 and J1857+0943). This approximation makes the S/N calculations of Siemens et al. (2013) optimistic; a demonstration of a similar optimism can be found in Yardley et al. (2011). Nonetheless, Siemens et al. (2013) predict that the NANOGrav PTA will, with 90% confidence, be able to detect a GWB from binary SMBHs with $A_{\rm yr} = 10^{-15}$ with $S/N = 3$ by the year 2021 with pessimistic assumptions about the level of intrinsic red timing noise for each pulsar. They expect NANOGrav to start timing three new pulsars each year with white noise levels at the median of the current sample, and assume 20 observations of each pulsar per year. While this calculation is likely optimistic, it does not account for observations by the EPTA and the PPTA, or for future observing hardware improvements. I hence assume in this thesis that a PTA detection of a GWB of the form in Equation 4.1 with $A_{\rm yr} = 10^{-15}$ is likely to be possible around the end of this decade.

4.2 Constraints on GWs from individual binary SMBHs

Z14 used the entire PPTA DR1 data set to search for GWs from individual binary SMBHs across the entire sky. All binary SMBHs were assumed to be in circular orbits that underwent insignificant evolution over the data span, which implies from Equations 3.21 and 3.22 that only sinusoidal GW signals were searched for. Equations 3.63 and 3.54 further imply that such signals will produce sinusoidal ToA variations. Z14 parameterised the amplitude of GWs emitted by a circular binary SMBH as

$$h_g = \frac{2(GM_C)^{5/3}}{c^4 D_L}(\pi f)^{2/3}. \tag{4.7}$$

This parameterisation assumes that the cosmological redshift has a negligible effect, which is valid for the nearby binary SMBHs that the Z14 search was sensitive to.

Both Z14 and W15 utilised fits for time series $A_+(t)$ and $A_\times(t)$ at many sky positions for their GW searches. These time series represent the magnitudes of

the timing residuals common to all pulsars with a quadrupolar spatial signature in the + and × GW polarisations respectively for a given sky position. Searching for GW signals in the $A_+(t)$ and $A_\times(t)$ time series, rather than in the raw data, greatly reduced the computational expense of the problem. $A_+(t)$ and $A_\times(t)$ were fit by Z14 at uniform intervals of 30 days using a global generalised least squares fit to the entire PPTA DR data set for 1000 sky positions. For each of 141 frequency channels between 2.5×10^{-9} Hz and 1.8×10^{-7} Hz, Z14 then measured the maximum-likelihood estimates of the complex Fourier coefficients in these time series, and used a likelihood ratio detection statistic to consider whether any of the measurements were significant.

Upon finding no significant detections, Z14 calculated a 95% confidence all-sky upper limit on the value of h_g at multiple frequencies. This was accomplished by performing numerous simulations of the DR1 data set with the injected effects of GWs from a single binary SMBH with arbitrary inclination, orbital phase, orientation of the line of nodes and sky position, but with fixed values of h_g and the GW radiation frequency. The value of h_g was adjusted until 95% of the simulations had a detection statistic greater than the measured one; this value was taken as the upper limit at the given frequency. An all-sky sensitivity curve was also produced, where now the simulated values of the detection statistic were compared with a threshold detection statistic with a single-trial false alarm probability of 10^{-4}. The upper limit and sensitivity curves are shown in Figure 4.4.

The sensitivity of an idealised PTA to individual binary SMBHs was investigated by Ellis et al. (2012). Sensitivity curves similar to that in Figure 4.4 are shown in Figure 4.5 for simulated 17- and 100-pulsar PTA data sets with bi-weekly observations over 5 yr. The timing noise in the simulated data sets was assumed to be white, with 100 ns rms for all pulsars. The losses in sensitivity at GW frequencies of $(1\,\mathrm{yr})^{-1}$ and $(0.5\,\mathrm{yr})^{-1}$, also seen in the PPTA DR1 sensitivity curve in Figure 4.4, are caused by the necessity of fitting for pulsar positions and parallaxes. The 'valleys' in the sensitivity curves at a frequency of $\sim 10^{-8}$ Hz are caused by the fitting of pulsar spin frequencies and frequency derivatives (see also Moore et al. 2014), which reduces sensitivity at low frequencies. It is important to note that the valley frequency is greater than the inverse data span of 6.3×10^{-9} Hz.

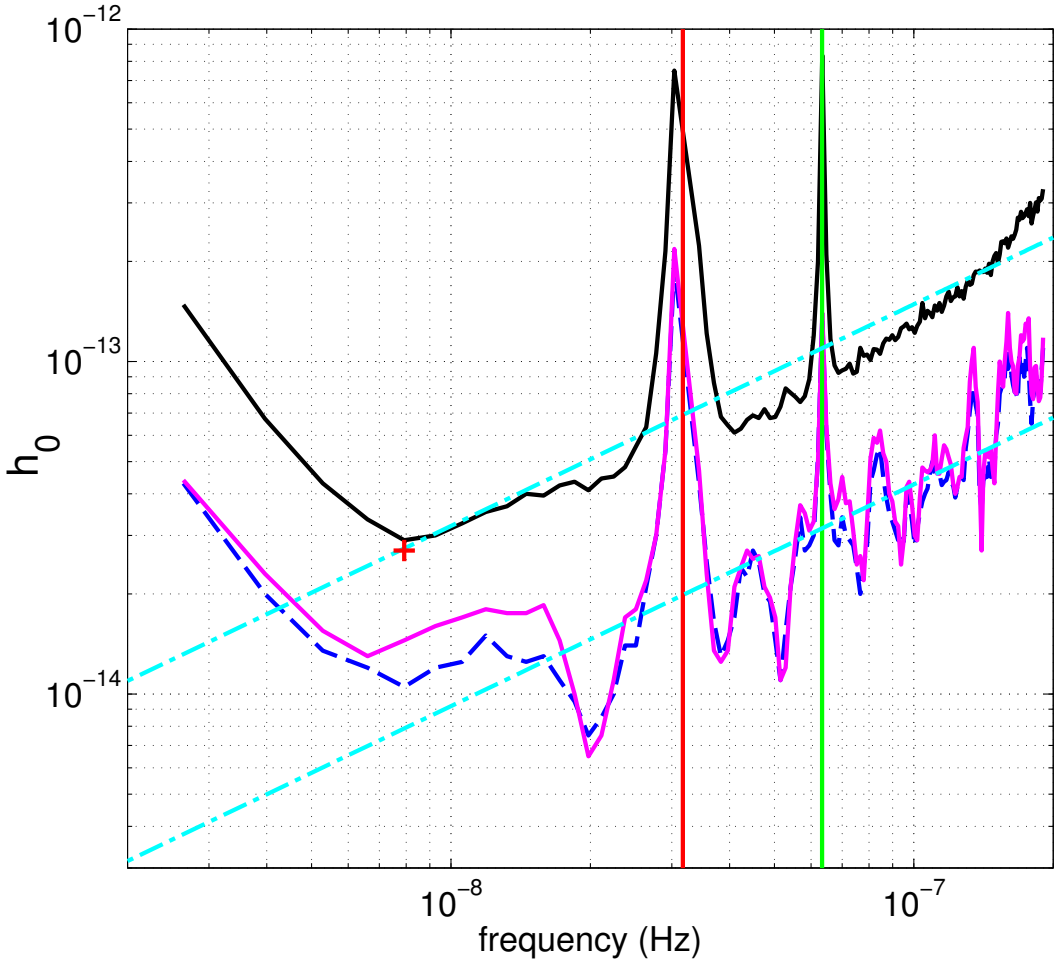

Figure 4.4: All-sky 95% confidence upper limits on h_g as a function of the received GW frequency, with results for simulations with real (blue dashed curve) and simulated noise (pink curve) both shown. The solid black curve shows the PPTA DR1 sensitivity to GWs from individual binary SMBHs. The cyan dot-dashed lines show values of h_g for binaries with $M_C = 10^{10} M_\odot$ and $D_L = 400\,\mathrm{Mpc}$ (upper) and with $M_C = 10^9 M_\odot$ and $D_L = 30\,\mathrm{Mpc}$ (lower). This figure is reproduced from Zhu et al. (2014).

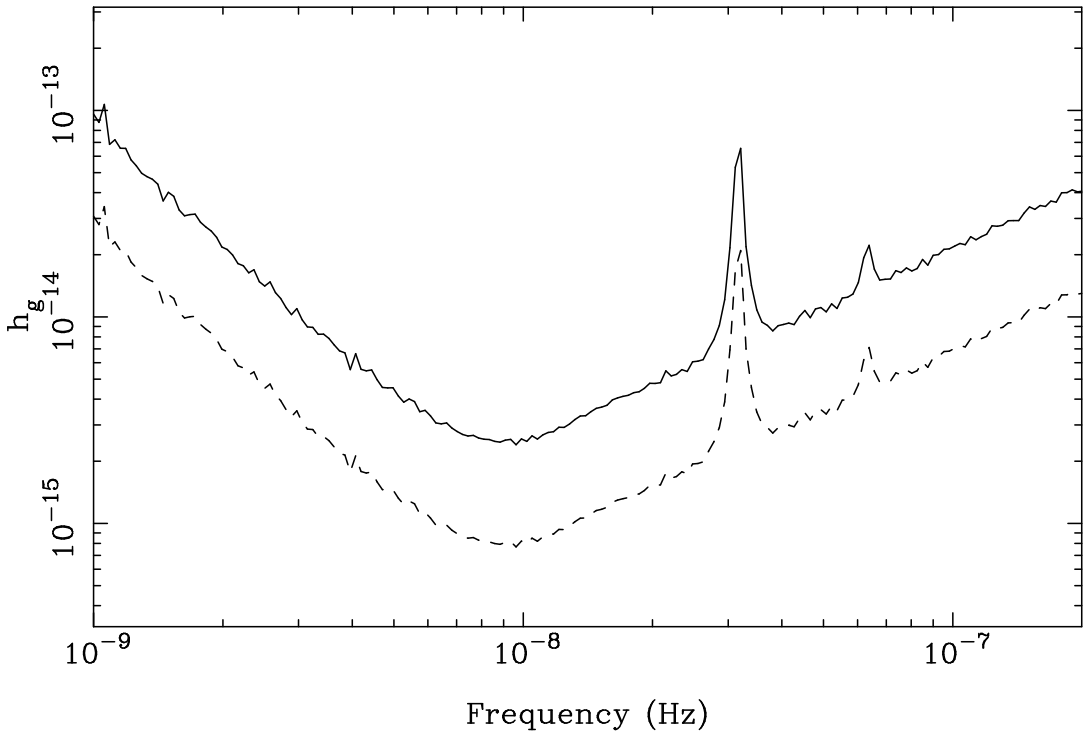

Figure 4.5: Predicted sensitivities to GWs from individual circular binary SMBHs for PTAs consisting of 17 pulsars timed with 100 ns precision (solid curve) and 100 pulsars timed with 100 ns precision (dashed curve) from Ellis et al. (2012). Data for these curves were kindly provided by J. Ellis. In both cases, uniform sky distributions and bi-weekly observations for 5 yr are assumed. The sensitivity values correspond to GW amplitudes, h_g, that are expected to be detected, with 95% confidence, at a false alarm probability of 10^{-4}.

4.3 All-sky limits on the rate of GW memory bursts

The W15 search for GW memory bursts was conducted using essentially the same fitted $A_+(t)$ and $A_\times(t)$ time series for different sky positions as the Z14 search. As is evident from Equation 3.63, a step change in the spatial part of the metric caused by a memory burst will cause a step change in the observed pulsar rotation frequency, which, from Equation 3.54, will result in ToA variations that increase in magnitude linearly with time. Hence, the effect of a memory burst on pulsar timing data is analogous to the effects of a rotational glitch, commonly observed in normal pulsars (e.g., Radhakrishnan & Manchester 1969) and observed once in a millisecond pulsar (Cognard & Backer 2004). However, a glitch caused by a GW memory burst will appear simultaneously for different pulsars with a size which has a quadrupolar dependence on the pulsar positions. The $A_+(t)$ and $A_\times(t)$ time series, for the optimal position, will therefore exhibit a linear ramp beginning when the memory burst propagates past the Earth. Of course, the ToA variations contributed by the 'pulsar terms' (see Equation 3.62) will be identical to the ToA variations contributed by the 'Earth term', but the pulsar term contributions will occur at different times for different pulsars because of differing pulsar distances and positions.

W15 were able to constrain the all-sky rates of GW memory bursts of different amplitudes, $h_{\rm mem}$. For example, they showed that bursts with $h_{\rm mem} = 6 \times 10^{-14}$ occur less than once per year. For a binary system consisting of equal-mass SMBHs with masses $M_{\rm BH}$ at a comoving distance D,

$$h_{\rm mem} \sim 2 \times 10^{-16} \frac{M_{\rm BH}}{10^8 M_\odot} \frac{1\,{\rm Gpc}}{D}. \tag{4.8}$$

Hence, $h_{\rm mem} = 6 \times 10^{-14}$ corresponds, for example, to the coalescence of two $3 \times 10^9 M_\odot$ SMBHs at a distance of 100 Mpc. This is likely to be quite rare. van Haasteren & Levin (2010) derived an expression for the detectable memory burst amplitude for an idealised PTA consisting of $N_{\rm psr}$ pulsars timed at cadence of ΔT over a time span $T_{\rm span}$, each with white timing noise with standard deviations $\sigma_{\rm psr}$. This expression was

$$h_{\rm mem}^{\rm detectable} = 95\sigma_{\rm psr}\sqrt{\frac{\Delta T}{N_{\rm psr} T_{\rm span}}} \sim 5 \times 10^{-16} \frac{\sigma_{\rm psr}}{100\,{\rm ns}} \left(\frac{\Delta T}{2\,{\rm weeks}} \frac{100}{N_{\rm psr}} \frac{5\,{\rm yr}}{T_{\rm span}} \right)^{1/2}, \tag{4.9}$$

implying that the 100-pulsar PTA with 100 ns timing residuals considered above could result in the detection of a memory burst with $h_{\rm mem} = 5 \times 10^{-16}$, if one occurred during the 5 yr observing time span.

Chapter 5

The effects of GWs from binary SMBHs on PTA data

I investigate the effects of gravitational waves (GWs) from the cosmological population of binary super-massive black holes (SMBHs) on pulsar timing array datasets. I construct a distribution describing the binary SMBH population from an existing semi-analytic galaxy formation model. Using realisations of the binary SMBH population generated from this distribution, I simulate pulsar timing datasets with GW-induced variations. I find that the statistics of these variations do not correspond to an isotropic GW background. The "Hellings & Downs" correlations between simulated datasets for different pulsars are recovered on average, although the scatter of the correlation estimates is greater than expected for an isotropic, stochastic GW background. These results are attributable to the fact that just a few GW sources dominate the GW-induced variations in every Fourier frequency bin of a 5-year dataset. Previous constraints on the amplitude of the GW signal from binary SMBHs are likely biased. Individual binary systems may be detectable in 5-year pulsar timing array datasets where the noise is dominated by GW-induced variations. Searches for GWs in pulsar timing array data therefore need to account for the effects of individual sources of GWs. This Chapter describes my original work, and the results have largely been published (Ravi et al. 2012; Shannon et al. 2013).

5.1 Introduction

Detecting and performing science with GWs is currently a major goal of experimental astrophysics. PTAs are sensitive to GWs in the frequency band $10^{-9} - 10^{-7}$ Hz, and are hence complementary to other ground-based and proposed space-based GW detection experiments. In this frequency band, the most promising astrophysical sources of GWs are binary super-massive black holes SMBHs.

It is likely that there exists a large population of binary SMBHs at various stages of coalescence in the cores of galaxies that have recently merged with other galaxies. The final stages of SMBH-SMBH coalescence are driven by losses of energy and angular momentum to GWs, primarily emitted in the PTA frequency band. Various works have predicted the average spectrum of the GW strain amplitude from the cosmic population of binary SMBHs (Rajagopal & Romani 1995; Jaffe & Backer 2003; Wyithe & Loeb 2003a; Enoki et al. 2004; Sesana et al. 2008b). Under the assumption that all binary systems are in circular orbits evolving only through GW emission, this characteristic strain spectrum, $h_c(f)$, takes the form of Equation 3.46 with exponent $\alpha = -2/3$, i.e.,

$$h_c(f) = A_{\rm yr} \left(\frac{f}{f_{\rm yr}} \right)^{-2/3} \tag{5.1}$$

The combined GW signal from binary SMBHs is widely assumed to form an isotropic, stochastic GW background (GWB; e.g., Hobbs et al. 2009). The value of $A_{\rm yr}$, which is the characteristic strain amplitude at a frequency of $f_{\rm yr}$ (see Equation 3.46), is used to specify the amplitude of the background. $A_{\rm yr}$ is always calculated by summing over the GW signals from all observable binary SMBHs in a specific model. Most immediately prior to this thesis, predictions for $A_{\rm yr}$ were made by Sesana et al. (2008b), who found a likely range of 10^{-16} to 2.5×10^{-15}. This range of predictions was derived by considering a variety of SMBH seeding, accretion and feedback scenarios, as well as uncertainties in the galaxy merger rate and in the SMBH mass function.

PTA projects are based on observations of periodic pulses of radio emission from pulsars using large radio telescopes. These observations lead to measurements of pulse ToAs at the observatories, which can be modelled using combinations of deterministic and stochastic processes. GWs incident on the Earth (and on the pulsars) cause shifts in the measured pulse frequencies of the pulsars (Sazhin 1978; Detweiler 1979). For a pulsar, indexed by p, with intrinsic rotation frequency ν_p,

consider a GW-induced shift, $\Delta\nu(t, \mathbf{r_p})$, to this frequency. This shift is a function both of time, t, and the Earth-pulsar direction vector, $\mathbf{r_p}$. The resulting discrete time-series of GW-induced variations to the ToAs, δ_i^p (the i subscript indicates that δ_i^p is sampled at times t_i), is given by the time-integral of $\Delta\nu(t, \mathbf{r_p})/\nu_p$ in Equation 3.54. For any stochastic, isotropic GWB, the expected value of the normalised, zero-lag cross-correlation between the δ_i^p time-series for each pulsar pair is expressed in terms of the angular separations between the pulsars as the Hellings & Downs curve (Hellings & Downs 1983; Jenet et al. 2005), given in Equation 3.66.

The Hellings & Downs curve ceases to describe the expected values of the cross-correlation coefficients if any part of the GW signal is *determined*. For example, if it is known that there is a region of the sky that is devoid of GWs, or if the GW polarisations for different regions of the sky are not truly random, correlations that are different from the the Hellings & Downs curve will be expected. Measurements of correlations between pulsar timing datasets that are attributable to the effects of GWs are necessary for the detection of GWs with PTAs (Jenet et al. 2005; Yardley et al. 2011; van Haasteren et al. 2011; Demorest et al. 2013).

The prospect of detecting or constraining the amplitude of a background of GWs from binary SMBHs has been the primary rationale for the development of the PTA concept. Various works have placed upper bounds on the value of $A_{\rm yr}$ for a GWB with the characteristic strain spectral form of Equation 5.1 (Jenet et al. 2006; van Haasteren et al. 2011; Demorest et al. 2013; Shannon et al. 2013). However, all analysis methods developed to study the combined GW signal from binary SMBHs with PTAs assume that the δ_i^p time-series for multiple pulsars can be described as a specific stochastic process (see, e.g., §4.1). I describe the exact nature of this assumption in §5.2.

In this Chapter, I elucidate the statistical nature of the ToA variations induced by GWs from the cosmological population of binary SMBHs. I accomplish this by modelling the GW signals from all binary SMBHs, and by simulating realisations of δ_i^p corresponding to realisations of the collation of these GW signals. This study is critical to the validity of interpreting published upper limits on $A_{\rm yr}$ as representative of limits on the mean characteristic strain spectrum of GWs predicted to arise from binary SMBHs. That is, if assumptions for the statistics of GW-induced ToA variations do not match what is actually expected for a population of binary SMBHs, one cannot directly compare upper limits on $A_{\rm yr}$ to models for the binary SMBH

Table 5.1: List of symbols.

Symbol	Section	Description
δ_i^p	5.1	GW-induced ToA variation for pulsar p at time t_i.
ρ_{pq}	3.2.2	Expected zero-lag normalised cross-correlation between δ_i^p and δ_i^q time-series.
$S_g(f)$	3.2.2	Expected PSD of δ_i^p time-series.
$\tilde{S}_k^p$	5.2	Periodogram estimator of $S_g(f)$ at frequency f_k.
h_0	5.3.1	GW strain amplitude divided by frequency dependence.
Φ	5.3.1	Binned distribution of binary SMBHs derived from realisations of the Millennium and Millennium-II coalescence lists.
$\bar{\Phi}$	5.3.1	Average of 1000 realisations of Φ.
$\Phi_{\rm fit}$	5.3.1	Analytic fit to $\bar{\Phi}$.
$S_{\rm g,\,fit}(f)$	5.3.2	$S_g(f)$ derived in terms of $\Phi_{\rm fit}$.
$h_{\rm c,\,fit}(f)$	5.3.2	Expected GW characteristic strain spectrum derived in terms of $\Phi_{\rm fit}$.
W_i	5.4	1 ns rms ToA variation at time t_i.
D_i^p	5.4	Sum of W_i and δ_i^p.
$S(f)$	5.4	Expected PSD of D_i^p time-series.
$\tilde{\psi}_k^p$	5.4	Periodogram estimator of $S(f)$ at frequency f_k.
$\tilde{\rho}_{pq}$	5.5	Estimator of ρ_{pq}.

population. These results are also important for the optimisation of GW detection techniques with PTAs. In §5.3, I outline the method of simulating pulsar timing datasets including GWs from the predicted population of binary SMBHs. My analysis and results are presented in §5.4 and §5.5, and I discuss the interpretation and implications of these results in §5.6. I summarise my conclusions in §5.7.

In the following, I assume a ΛCDM concordance cosmology based on a combined analysis of the first-year WMAP data release (WMAP1; Spergel et al. 2003) and the 2dF Galaxy Redshift Survey (Colless et al. 2001), with $\Omega_M = 0.25$, $\Omega_b = 0.045$, $\Omega_\Lambda = 0.75$, $\sigma_8 = 0.9$ and $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$. Although these parameter values have since been superseded by more recent observations, I adopt them in order to remain consistent with the model I use for the binary SMBH population (Guo et al. 2011). A list of important symbols in this Chapter is shown in Table 5.1, along with the sections of the text in which they are introduced.

5.2 Gaussian GW-induced ToA variations

ToA variations induced by GWs from binary SMBHs (δ_i^p) are commonly modelled among the PTA community as a wide-sense stationary random Gaussian process. This is based on the hypothesis that many GW sources forming a GWB contribute equally to the ToA variations, resulting in a statistical process governed by the central limit theorem. While the nature of the random Gaussian model for δ_i^p has been extensively described elsewhere (e.g., van Haasteren et al. 2009), I summarise it here for completeness.

The key property of a random Gaussian process is that a linear combination of samples will have a joint normal distribution function. Different samples need not be statistically independent. The distribution of samples from a (zero-mean) random Gaussian process is characterised only by the covariance matrix of the samples. Consider a vector, $\mathbf{R_p}$, containing n samples of δ_i^p. That is,

$$\mathbf{R_p} = \begin{pmatrix} \delta_0^p \\ \delta_1^p \\ \dots \\ \delta_{n-1}^p \end{pmatrix}. \tag{5.2}$$

Let $\mathbf{R_q}$ be another vector defined similarly to $\mathbf{R_p}$, corresponding to a pulsar q, containing n simultaneously-obtained samples of δ_i^q. Under the random Gaussian assumption, the joint probability distribution of the samples in $\mathbf{R_p}$ and $\mathbf{R_q}$, which I denote P_{pq}, is given by

$$P_{pq} = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C_{pq}})}} e^{-\frac{1}{2}\mathbf{R_p}^T \mathbf{C_{pq}} \mathbf{R_q}}. \tag{5.3}$$

Here, $\mathbf{C_{pq}}$ is an $n \times n$ matrix containing the covariances between the samples of δ_i^p and δ_i^q; that is, element ij of $\mathbf{C_{pq}}$ is given by the covariance between δ_i^p and δ_j^q. As the Gaussian process is wide-sense stationary, each element ij of C_{pq} depends only on the time difference $\tau_{ij} = |t_i - t_j|$ between samples i and j for pulsar p and q respectively. Elements of $\mathbf{C_{pq}}$ are sampled from a covariance function, $c_{pq}(\tau)$, between the GW-induced ToA variations for pulsars p and q. This covariance function is defined by the inverse Fourier transform of the one-sided power spectral density (PSD), $S_g(f)$,

of GW-induced ToA variations for a given pulsar:

$$c_{pq}(\tau) = \rho_{pq}\text{Real}[\mathcal{F}^{-1}(S_g(f))]. \tag{5.4}$$

Here, $\mathcal{F}$ denotes a complex Fourier transform, and τ is a time-lag. The PSDs of the GW-induced ToA variations for all pulsars are equivalent, and given by Equation 3.65 for a GW signal with the expected characteristic strain spectrum $h_c(f)$.

The above discussion applies equivalently if pulsar p and pulsar q are the same pulsar, or if they are different pulsars. The Hellings & Downs factor ρ_{pq}, defined in Equation 3.66, accounts for the correlation between GW-induced TOA variations for different pulsars. If $h_c(f)$ has the power-law form in Equation 5.1, then $S_g(f) \propto f^{-13/3}$. The GW-induced ToA variations for each pulsar will therefore be a "red" process; that is, the variance of a GW-induced ToA variation time-series will be increased for longer time spans. In this work, I only consider time-series δ_i^p with finite lengths, T_{obs}.

Here, I am interested in comparing a new model for δ_i^p with the random Gaussian model described in this section. To this end, it is necessary to be able to simulate realisations of δ_i^p as a random Gaussian process. Multiple PTA groups test their data analysis algorithms by simulating realisations of δ_i^p using the GWbkgrd plugin (Hobbs et al. 2009) to the TEMPO2 pulsar timing package (Hobbs et al. 2006). In this plugin, a number of GW oscillators, N_{T2}, are simulated between GW frequencies f_{lo} and f_{hi}, with the normally distributed + and × GW polarisation amplitudes set to be purely real with zero mean, variance

$$\sigma_{T2}^2 = \sqrt{\frac{\ln(f_{\text{hi}}/f_{\text{lo}})}{N_{T2}}} h_c(f), \tag{5.5}$$

and frequency probability distribution, dP/df, given by

$$\frac{dP}{df} = \frac{1}{\ln(f_{\text{hi}}/f_{\text{lo}})} f^{-1}. \tag{5.6}$$

ToA variations calculated for a given pulsar p at different times t_i corresponding to GWs from each of these oscillators are summed to produce a realisation of the δ_i^p time-series. The frequency limits f_{lo} and f_{hi} are generally chosen respectively to be much less than the T_{obs}^{-1} and much greater than the Nyquist frequency corresponding

to the minimum sampling interval.

I make a distinction between the expected PSD of ToA variations induced by GWs from binary SMBHs, as defined in Equation 3.65, and estimates of this PSD based on realisations of the ToA variations. A commonly-used non-parametric, unbiased estimator of the PSD of a time-series is the periodogram (Schuster 1898). The periodogram, $\tilde{S}_g^p$, of δ_i^p is defined as

$$\tilde{S}_k^p = \frac{2}{T_{\rm obs}} |{\rm DFT}[\delta_i^p]|^2, \tag{5.7}$$

where DFT denotes a discrete Fourier transform. I adopt the following standard definition for the DFT of n samples of δ_i^p:

$${\rm DFT}(f_k) = \sum_{m=0}^{n-1} \delta_m^p e^{-i2\pi mk/n} \frac{T_{\rm obs}}{n}, \tag{5.8}$$

where $i = \sqrt{-1}$ in this case. The DFT is evaluated for frequencies

$$f_k = (k+1)\frac{1}{T_{\rm obs}},\ 0 \leq k < \frac{T_{\rm obs}}{2T_{\rm samp}}, \tag{5.9}$$

where $T_{\rm samp}$ is the interval (assumed to be constant) between samples of δ_i^p. Throughout this work, I estimate the PSD, $S_g(f)$, of realisations of δ_i^p by evaluating $\tilde{S}_k^p$.

While the GWbkgrd plugin does not explicitly generate random Gaussian realisations of δ_i^p by construction, I and others (van Haasteren et al. 2011; Demorest et al. 2013) have checked that it approximates a random Gaussian process well. A commonly-used method of producing a realisation of a time series corresponding to a time-correlated random Gaussian process is to first produce a realisation of its complex DFT with normally distributed real and imaginary components at each frequency, and then perform an inverse DFT on this (e.g., Coles & Filice 1984). This technique has the advantage of being able to quickly produce a very long complex DFT, corresponding to a very long time series from which separate sections of the desired length can be extracted. I performed various comparisons between realisations of GW-induced ToA variations made using this technique and the GWbkgrd plugin, including the distributions of samples at specific times and the distributions of PSD estimates $\tilde{S}_k^p$, and found no significant differences.

5.3 Simulating GW-induced pulsar ToA variations

In this section, I describe a new method of simulating ToA variations caused by GWs from the predicted population of binary SMBHs. Various works have presented models for the cosmic demographics of binary SMBHs (Dotti et al. 2012, and references therein). Here, I use results from an existing semi-analytic model for galaxy and SMBH formation, which includes analytic prescriptions for baryon physics applied to dark matter halo merger tree catalogues from the Millennium (Springel et al. 2005) and the Millennium-II (Boylan-Kolchin et al. 2009) simulations (Guo et al. 2011, hereafter G11 in this Chapter). The Millennium and Millennium-II simulations both follow the evolution of dark matter structures, using the same physical prescriptions and number of particles. The Millennium-II simulation was however carried out in a comoving cubic volume with one-fifth the side length as that of the Millennium simulation, with the aim of resolving smaller-scale dark matter structures than the Millennium simulation.[1] Together, these simulations resolve dark matter haloes corresponding to the observed galaxy population, from dwarf galaxies to the largest-mass early-type galaxies.

The G11 model is the latest in a series (Springel et al. 2005; Croton et al. 2006; De Lucia & Blaizot 2007) of semi-analytic prescriptions applied to the Millennium simulations. A host of observables of galaxies at low redshifts are reproduced, along with the redshift-evolution of the quasar population and star formation. Of most relevance here is that the model also traces the SMBH population, reproducing the $z = 0$ SMBH-galaxy scaling relations in their slopes, normalisations and scatters, as well as the inferred SMBH mass function (Marulli et al. 2008). The G11 model treatment of the growth of SMBHs is exactly the same as in the Croton et al. (2006) model, as described in §2.3 by Equations 2.29 and 2.30.

I base my description of the binary SMBH population emitting GWs on the prediction for the SMBH-SMBH coalescence rate from the G11 model. I fit an analytic function to the distribution of binary SMBHs, and randomly draw GW sources from this distribution to produce realisations of the GW sky corresponding to binary SMBHs. I then sum the effects of each GW source on simulated pulsar ToA datasets in order to analyse the GW-induced ToA variations.

[1]The Millennium-II simulation however does not reproduce larger-scale structures as well as the Millennium simulation.

My work is different from most previous attempts to model the GW signal from binary SMBHs. Initial attempts (e.g., Jaffe & Backer 2003) to predict the mean GW characteristic strain spectrum from binary SMBHs used empirical determinations of the galaxy merger rate and the SMBH mass function. Wyithe & Loeb (2003a) predicted the GW characteristic strain spectrum by analytically following the dark matter halo merger hierarchy in the EPS framework, and by deriving the SMBH coalescence rate by relating the SMBH masses to the halo circular velocities. Sesana et al. (2008b) considered the possible range of predictions of the characteristic strain spectrum, using Monte Carlo realisations of dark matter halo merger trees and various prescriptions for SMBH growth.

The key difference between the present work and previous calculations of the GW signal from binary SMBHs is that I am chiefly concerned with the statistics of δ_i^p. My approach to modelling the binary SMBH population is similar to Sesana et al. (2009) in the use of mock galaxy catalogues derived from analytic prescriptions applied to the Millennium simulations. However, whereas Sesana et al. (2009) modelled the SMBH population by using empirical SMBH-galaxy scaling relations combined with (earlier) mock catalogues, I utilise SMBHs modelled by G11 in a self-consistent framework which reproduces the relevant observables.

5.3.1 Modelling the distribution of binary SMBHs

As in the previous works discussed above, I considered all binary SMBHs to be in circular orbits, and used expressions for the resulting GW emission presented in §3.1.2. I briefly discuss the assumption of circular orbits in §5.6.3. The mock galaxy catalogues resulting from the G11 model are available online[2] (Lemson & Virgo Consortium 2006). The halo merger trees from the Millennium and Millennium-II simulations were evaluated at 60 logarithmically-spaced redshift "snapshots" between $z = 0$ and $z = 20$ (see Equation 2.17). I obtained the lists of SMBH-SMBH coalescence events within the comoving volume of each simulation by querying the online database. Redshifts at the (non-logarithmic) midpoints between the redshift snapshots were assigned to each event. I used these lists to fill bins of a distribution, Φ, of the number, N, of observable binary SMBHs per unit comoving

[2]http://www.mpa-garching.mpg.de/millennium/

volume per solid angle on the sky, given by

$$\Phi = \frac{dN}{dh_0} 4\pi \frac{d^2V_c}{d\Omega dz} \frac{dz}{dt} \frac{dt}{df} \tag{5.10}$$

where

$$h_0 = \left(\frac{(GM_C)^{5/3}}{c^4 D(z)} (\pi(1+z))^{2/3} \right)^2 = (\sqrt{\frac{5}{32}} h_s f^{-2/3})^2, \tag{5.11}$$

and $4\pi \frac{d^2V_c}{d\Omega dz}$ is the sky-integrated comoving volume shell between redshifts z and dz. Also, $\frac{dz}{dt} = H_0 \sqrt{\Omega_M (1+z)^3 + \Omega_\Lambda}$, and the derivative $\frac{dt}{df}$ was obtained from Equation 3.33. Φ is the predicted distribution of binary SMBHs in h_0 (which corresponds to the frequency-independent GW "power", or squared strain amplitude) and the observed GW frequency.

For chirp masses below $10^7 M_\odot$, the limited capability of the Millennium simulation to resolve low-mass haloes caused an under-prediction of the chirp mass function as compared to the Millennium-II simulation. In order to ensure a complete chirp mass function, I included binary SMBHs with $M_C > 10^7 M_\odot$ from the Millennium list of coalescence events, and binaries with $10^6 M_\odot < M_C \leq 10^7 M_\odot$ from the Millennium-II list.

Some degrees of randomisation in the coalescence lists were possible. First, in cases where more than two SMBHs coalesced to form a single SMBH between redshift snapshots, the merger order was not specified. In these instances I randomised over merger order. Second, a spherical comoving volume shell between any pair of redshifts less than ~ 0.09 could be contained within the simulation volume. Some Millennium redshift snapshots exist at $z < 0.09$, and the comoving volume shells between redshifts corresponding to these snapshots enclose some SMBH-SMBH coalescence events in the G11 model. An observer located at the centre of the Millennium simulation volume would observe only a fraction of the total list of events in the G11 model at $z < 0.09$, and an observer located elsewhere in the volume would observe a different selection of events. This is not the case for the Millennium-II simulation, where the volume was too small to enclose any comoving volume shells between redshift snapshots. For each realisation of the Millennium (but not the Millennium-II) coalescence list, I therefore specified randomly-placed spherical shells within the simulation box to select binary SMBHs at these redshifts. For coalescence events at $0.09 < z < 0.19$, the corresponding comoving volume shells between redshift snap-

shots were smaller than the Millennium simulation volume, though not enclosed by it. For these coalescence events, I randomly included binaries in the Millennium coalescence list according to probabilities given by the ratios between the volumes of the comoving shells and the Millennium simulation volume. I used 1000 realisations of the Millennium and Millennium-II coalescence lists to form realisations of the binary SMBH distribution Φ.

In generating realisations of the distribution Φ, I assumed that every SMBH-SMBH coalescence event in the G11 model catalogues was the result of a binary SMBH system that had decayed through GW emission. The G11 model included the assumption that, upon the merger of two galaxies with central SMBHs, the SMBHs coalesced in every case, before accretion onto the newly-formed SMBH.[3] I note that SMBHs with masses as low as $10^3 M_\odot$ were present in the G11 model catalogues, but were not included in the Φ distributions. I verified that relaxing the lower cutoff on the SMBH masses in the Φ distributions from $10^6 M_\odot$ to $10^3 M_\odot$ did not significantly modify the total signal.

The 1000 realisations of Φ were averaged to form a distribution $\bar{\Phi}$. I fitted $\bar{\Phi}$ with an analytic function which could be used to generate random realisations of the observable binary SMBH population. I did not use realisations of Φ as realisations of the binary SMBH population because the Φ-distributions were binned for computational purposes. A four-parameter function,

$$\Phi_{\rm fit} = n \left(\frac{h_0}{p_h}\right)^{\alpha} \left(1 + \frac{h_0}{p_h}\right)^{\beta} f^{-11/3}, \tag{5.12}$$

with free parameters n, p_h, α and β, was found to fit $\bar{\Phi}$ well. I performed the fit on the logarithm of the data in order to approximate least-squares assumptions in the fitting procedure. The best-fit parameters are given in Table 5.2. The frequency-exponent was held fixed at $-11/3$, as predicted by Equations 5.10 and 3.33. See Appendix A (§5.8.2) for more discussions on fitting $\bar{\Phi}$.

[3]There are various mechanisms by which extreme mass ratio binary SMBH systems and triple or higher-order systems can avoid coalescence (e.g., Volonteri et al. 2003).

5.3.2 Realisations of pulsar ToAs with GW-induced variations

I used a rejection sampling algorithm with a logarithmic envelope to draw sets of binary SMBHs from the distribution $\Phi_{\rm fit}$. I simulated corresponding ToA variation time-series, δ_i^p, by summing the contributions caused by GWs from each individual binary. Details of the method used to calculate these contributions are presented in Hobbs et al. (2009). For each binary, I randomised over the right ascension and declination, the orbital inclination angle, the orientation of the line of nodes, and the orbital phase angle at the line of nodes. A publicly-available TEMPO2 library, "vikramSim", was written to perform this simulation.

For most of the present work, I did not use TEMPO2 to fit timing model parameters. Instead, I made use of the tools available for spectral analysis of timing residuals. In the simulations, the "timing residuals" corresponded exactly to δ_i^p given the absence of timing model fitting.

There is an important distinction between the δ_i^p time-series and the timing residuals resulting from analyses of observed ToA datasets. Consider a set of observed ToAs that exactly match a particular timing model, except for the addition of GW-induced variations (a δ_i^p time-series). Given that an observer does not actually possess any prior knowledge of the timing model parameters, the observer will fit the model parameters to the ToAs. The resulting timing residuals will not be equivalent to δ_i^p. This is because the δ_i^p variations in the ToAs can alter the apparent pulsar timing parameters. For example, the presence of a δ_i^p time-series consisting of a sinusoidal signal with a period of one year will alter the apparent pulsar position. This point is illustrated in Figure 3.2 and the related discussion.

In order to investigate the statistics of δ_i^p given the model for the binary SMBH population, I first used TEMPO2 to generate 500 ToAs spanning 5 years exactly corresponding to the PSR J0437−4715 timing model (Manchester et al. 2013). I then added realisations of the δ_i^p time-series evaluated at the observed ToAs to these datasets, that is, with $T_{\rm obs} = 5\,{\rm yr}$ and $T_{\rm samp} = 0.01\,{\rm yr}$. The pulsar distance was set to 1 kpc, and the position was held fixed for all simulations. As the binary SMBHs used to produce realisations of δ_i^p had randomised positions and orientations, allowing the pulsar position to vary between realisations would not alter the results. The results presented here are not dependent on the timing model used or on the pulsar distance

Table 5.2: Best-fit parameter values of $\Phi_{\rm fit}$.

Parameter	Value
n	2087 ± 365
p_h	$4.878{\times}10^{-23}{\pm}4.45 \times 10^{-24}$
α	$-1.72249{\pm}0.00064$
β	$-0.3473{\pm}0.0046$

from the Earth. I also added Gaussian white noise variations with 1 ns rms to the ToAs. This white noise component is much smaller than is usually observed in ToA datasets, but was necessary to smooth over machine precision errors.

I included GW sources between 10^{-9} Hz and 10^{-6} Hz in the simulations of δ_i^p. The lower frequency cutoff was chosen to be less than one fifth of $f_0 = (5\,\text{years})^{-1}$. The upper frequency cutoff was chosen to be greater than $f_{249} = (0.02\,\text{years})^{-1}$. I assumed, after previous works, that all GW sources between these frequency cutoffs are non-evolving over a 5-year timespan, i.e., they do not evolve in frequency by more than $(5\,\text{years})^{-1}$. The time, $t_{\rm evo}$, taken by a binary SMBH to double its emitted GW frequency, f_r, can be derived by adapting Equation 3.33:

$$t_{\rm evo} = 37492\,\text{yr}\left(\frac{M_C}{10^9 M_\odot}\right)^{-5/3}\left(\frac{f_r}{10^{-8}\,\text{Hz}}\right)^{-8/3}. \tag{5.13}$$

This justifies my assumption.

While the ultimate upper bound on the h_0-values of sources, $h_{0,\,\rm max}(f)$, was set by the last stable orbit of binary SMBHs, I more crudely identified this bound by fitting a power law to the high-h_0 edge of the $\bar{\Phi}$ distribution. This edge was defined somewhat arbitrarily as the h_0-value at each frequency above which the GW source distributions were poorly defined, with bins that were empty of counts. This choice is not important for the purposes of this Chapter; I use a better prescription in Appendix A, where I present a careful calculation of the GWB amplitude predicted by the G11 model. The lower bound on h_0, $h_{0,\,\rm min}$, was set by the lowest non-zero h_0-value in $\bar{\Phi}$. This value corresponds to a binary SMBH containing two $10^6 M_\odot$ components at $z \approx 6$. The distribution included more than 6.5×10^{18} GW sources within this $h_0 - f$ domain; the vast computational cost involved makes it impossible to simulate δ_i^p with this many sources. Fortunately, the shape of the $\Phi_{\rm fit}$ distribution

was such that, at a given frequency, the highest-h_0 sources contributed most to $S_g(f)$ (I return to this point below), defined in terms of $\Phi_{\rm fit}$ as:

$$S_{\rm g,\,fit}(f) = \frac{1}{12\pi^2 f^2} \int_{h_{0,\,\rm min}}^{h_{0,\,\rm max}(f)} \Phi_{\rm fit} h_s^2(f) dh_0, \tag{5.14}$$

The average characteristic strain spectrum derived from $\Phi_{\rm fit}$ is

$$h_{\rm c,\,fit}(f) = \left(f \int_{h_{0,\,\rm min}}^{h_{0,\,\rm max}(f)} \Phi_{\rm fit} h_s^2(f) dh_0 \right)^{1/2}. \tag{5.15}$$

I found a function, $\hat{h}_0(f)$, such that

$$0.9 S_{\rm g,\,fit}(f) = \int_{\hat{h}_0(f)}^{h_{0,\,\rm max}(f)} \frac{\Phi_{\rm fit} h_s^2(f)}{12\pi^2 f^2} dh_0 = \hat{S}_{\rm g,\,fit}(f). \tag{5.16}$$

Thus, the GW sources in the domain $\hat{h}_0(f) < h_0 < h_{0,\,\rm max}(f)$ contribute, on average, 90% of $S_g(f)$ at every frequency. Between 10^{-9} Hz and 10^{-6} Hz, this amounted to $\sim 4.5 \times 10^6$ sources. I refer to this $h_0 - f$ domain as the "90% domain". The 90% domain, along with $h_{0,\,\rm max}(f)$, $\hat{h}_0(f)$ and $h_{0,\,\rm min}$, is shown in Figure 5.1.

I approximated the total number of sources (6.5×10^{18}) in the $h_0 - f$ domain between $h_{o,\,min}$ and $h_{0,\,\rm max}$ and between 10^{-9} Hz and 10^{-6} Hz as constant. For a given realisation of δ_i^p, the actual number of sources in the 90% domain is governed by binomial statistics. I therefore drew a (binomial-)random number of sources from the 90% domain, and added contributions from each of them to each realisation of δ_i^p. I assumed that the sources remaining in the $\Phi_{\rm fit}$ distribution with $h_{0,\,\rm min} \leq h_0 \leq \hat{h}_0(f)$, contributing on average 10% to $S_g(f)$ at every frequency, resulted in a stochastic contribution to δ_i^p governed by the central limit theorem. I therefore simulated them using the method of simulating ToA variations corresponding to a GWB implemented in the TEMPO2 plugin GWbkgrd. I simulated $N_{T2} = 5 \times 10^4$ sources between 10^{-9} Hz and 10^{-6} Hz using the TEMPO2 method, with the characteristic strain spectrum given by $h_c(f) = \sqrt{12\pi^2 f^3 (S_{\rm g,\,fit}(f) - \hat{S}_{\rm g,\,fit}(f))}$. For each realisation of δ_i^p, I added contributions from the $\sim 4.5 \times 10^6$ GW sources drawn from the 90% domain of the $\Phi_{\rm fit}$ distribution, and from the 5×10^4 GW sources corresponding to the remaining (on average) 10% of $S_g(f)$ drawn using the TEMPO2 method. Shifts in the measured pulse frequencies caused by metric perturbations at both the

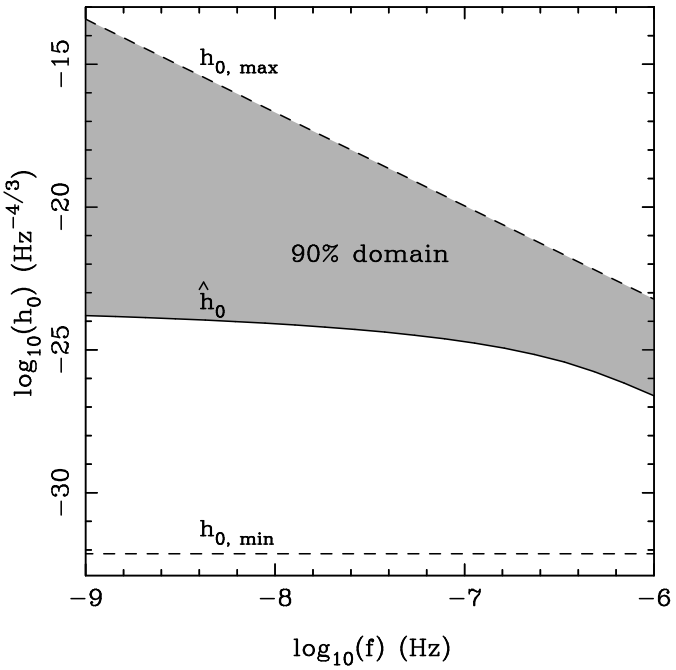

Figure 5.1: Illustration of the $h_0 - f$ domain constraints on the distribution $\Phi_{\rm fit}$. The upper and lower dashed lines represent $h_{0,\,\rm max}(f)$ and $h_{0,\,\rm min}$, as labelled, and the solid curve represents $\hat{h_0}(f)$. The shaded region is the "90% domain" from which binary SMBHs contributing, on average, 90% of the ToA variation PSD at every frequency were drawn.

Earth and the pulsar (i.e., the "Earth term" and the "pulsar term") were included in the simulations.

The top panel of Figure 5.2 shows $h_{\rm c,\,fit}(f)$ in the $0 \leq k < 100$ spectral bins. I also show a characteristic strain spectrum of the power-law form in Equation 5.1 with $A_{\rm yr} = 6.8 \times 10^{-16}$. This value of $A_{\rm yr}$ cannot, however, be taken to be the prediction from the G11 model, for reasons discussed further in Appendix A to this Chapter. The particular curvature in the $h_{\rm c,\,fit}$ curve, also predicted by Sesana et al. (2008b), is caused by the frequency-dependence of $h_{0,\,\rm max}(f)$, which represents the bound beyond which the binary SMBH distribution that I derive from the G11 model is sparse. Sesana et al. (2008b) fitted a broken power-law to realisations of the characteristic strain spectrum, accounting for various randomisations over the source population. In particular, Sesana et al. (2008b) randomised over the existence of "fractional" sources in every frequency bin of a fiducial dataset, and also excluded the strongest single source in every frequency bin in an attempt to isolate the background signal. The smaller number of sources per unit frequency at higher GW frequencies, combined with the greater contributions to the signal from the strongest single sources in frequency bins at higher frequencies, both resulted in the curved

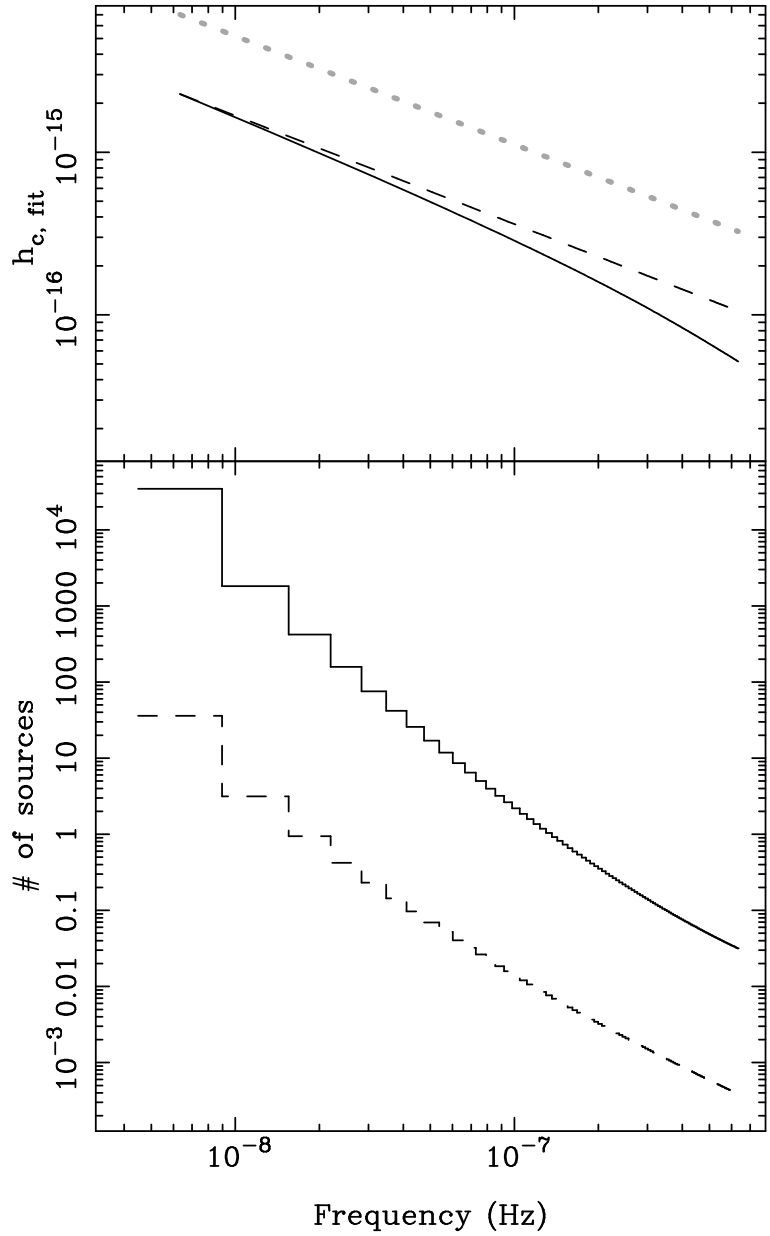

Figure 5.2: *Top:* The solid curve shows the mean characteristic strain spectrum, $h_{\mathrm{c,\,fit}}$, derived from the distribution Φ_{fit} in Equation 5.15. The dashed line shows a representative spectrum of the form in Equation 1, with $A_{\mathrm{yr}} = 6.8 \times 10^{-16}$. The values of both traces are equivalent at the lowest frequency. The dotted line shows a spectrum of the form in Equation 5.1 with $A_{\mathrm{yr}} = 2.4 \times 10^{-15}$, corresponding to the most recently published 95% confidence upper bound on A_{yr} (Shannon et al. 2013). *Bottom:* The numbers of GW sources that contribute 50% (dashed line) and 90% (solid line) of $S_{\mathrm{g,\,fit}}(f)$. The numbers are integrated over frequency bins of width (5 years)$^{-1}$ Hz.

characteristic strain spectra presented by Sesana et al. (2008b). Similarly, the increasing sparsity of the binary SMBH distribution with frequency in my case causes the curvature in the characteristic strain spectrum in Figure 5.2.

The bottom panel of Figure 5.2 shows the mean numbers of highest-h_0 sources that contribute 90% and 50% of $S_{\mathrm{g,\,fit}}(f)$ in these frequency bins. A small number of sources contribute a large fraction of $S_{\mathrm{g,\,fit}}(f)$ at every frequency. In the $k = 0$ frequency bin, the $\sim 3 \times 10^4$ highest-h_0 sources contribute on average 90% of $S_{\mathrm{g,\,fit}}(f)$, and only 30 sources on average contribute 50% of $S_{\mathrm{g,\,fit}}(f)$. At frequencies $f > 1.5 \times 10^{-7}\,\mathrm{Hz}$, the strongest source, on average, contributes more than 90% of $S_{\mathrm{g,\,fit}}(f)$ in each frequency bin. This is a consequence of the shallow power-law nature of the h_0-distribution of the GW sources in the Φ_{fit} distribution.

In this work, I compare the Millennium-based simulations of δ_i^p with simulations of δ_i^p created using the TEMPO2 method described in §5.2. To this end, I simulated ToAs as before, but added realisations of δ_i^p corresponding to 5×10^4 oscillators simulated using the TEMPO2 plugin GWbkgrd. These oscillators corresponded to a mean characteristic strain spectrum given by $h_{\mathrm{c,\,fit}}(f)$. I refer to this latter method of simulating δ_i^p as Case H09, after Hobbs et al. (2009). Simulations of δ_i^p using GW sources drawn from Φ_{fit} will be referred to as Case R12, as the present work was published in Ravi et al. (2012).

5.4 Fourier-spectral analysis, and results

In this section, I consider the differences between the cases in the distributions of the periodograms, $\tilde{S}_k^p$, evaluated for realisations of δ_i^p for a single pulsar. This is motivated by the results in Figure 5.2, in particular that the number of GW sources per spectral bin that contribute 90% of $S_{\mathrm{g,\,fit}}(f)$ varies significantly with frequency. The Case R12 simulations are intended to more accurately represent the effects of GWs from binary SMBHs on ToA datasets than the Case H09 simulations. Example realisations of 5-year δ_i^p time-series in both cases are shown in Figure 5.3. The time-series appear quite similar: realisations in both cases are dominated by low-frequency components. Values of up to $1\,\mu$s are also present in one realisation.

Instead of directly measuring $\tilde{S}_k^p$, the added white-noise component in the simulated ToAs required us to analyse the periodograms of a time-series, D_i^p, given by

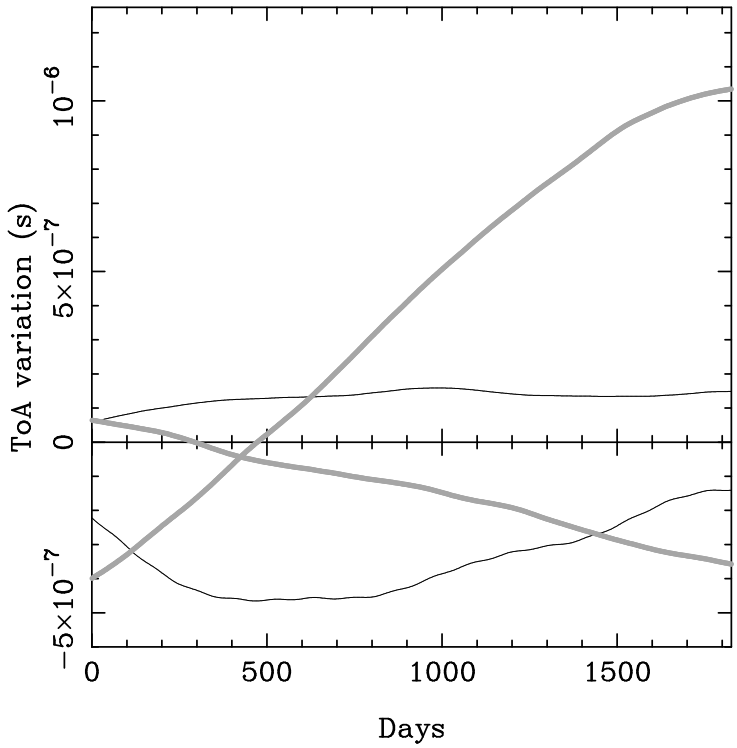

Figure 5.3: Example realisations of δ_i^p in Case H09 (thick grey lines) and Case R12 (thin black lines).

$$D_i^p = \delta_i^p + W_i, \tag{5.17}$$

where W_i is a time-series of Gaussian white 1 ns rms ToA variations as discussed above. The PSD of δ_i^p, $S_{\mathrm{g,\,fit}}(f)$, is significantly red, with a spectral index of $-13/3$ (see Equations 5.1 and 3.65), and is expected to dominate the PSD of W_i at low frequencies. I used the generalised least-squares algorithm described in Coles et al. (2011) to measure the periodograms, $\tilde{\psi}_k^p$, of realisations of D_i^p. This method requires knowledge of the auto-covariance function of D_i^p, which I obtained using the inverse DFT of the known PSD of D_i^p, $S(f)$, given by

$$S(f) = S_{\mathrm{g,\,fit}}(f) + \frac{2(1\,\mathrm{ns})^2}{250/(5\,\mathrm{years})}. \tag{5.18}$$

In the following, I consider the distributions of $\tilde{\psi}_k^p$ in the lower spectral bins, where $S(f) \approx S_{\mathrm{g,\,fit}}(f)$, to be approximately equivalent to the distributions of $\tilde{S}_k^p$.

I produced 1000 realisations of D_i^p in Case R12 and in Case H09, and measured $\tilde{\psi}_k^p$ for each realisation. In the top panel of Figure 5.4, I show the averages of $\tilde{\psi}_k^p$

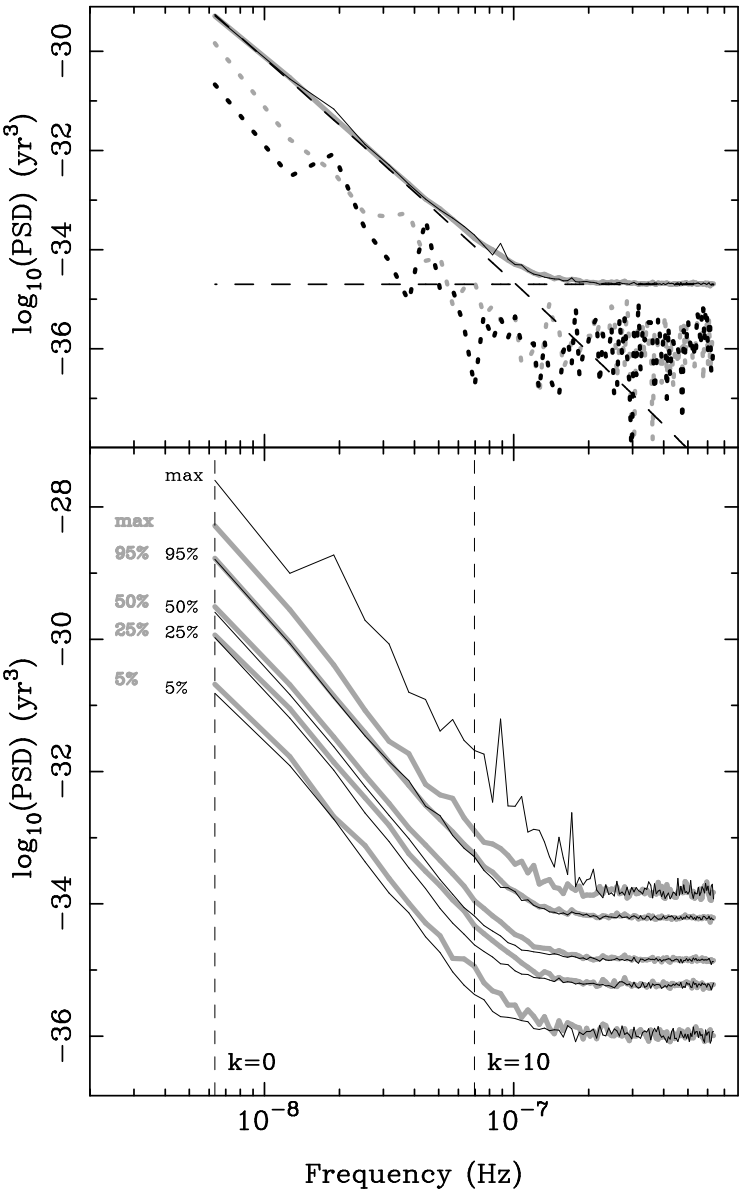

Figure 5.4: *Top:* The mean estimates ($\tilde{\psi}_k^p$) of the PSD of the simulated ToA variations (D_i^p) in Case R12 (thin solid black line) and in Case H09 (thick solid grey line). The predicted PSDs of δ_i^p ($S_{\mathrm{g,\,fit}}(f)$) and W_i are shown as sloped and horizontal dashed lines respectively. Randomly-chosen single measurements of $\tilde{\psi}_k^p$ in Case R12 and Case H09 are also shown, scaled down by a factor of 10, as black and grey dotted lines respectively. *Bottom:* The thin black and thick grey curves depict "percentile periodograms" of the distributions of Case R12 and Case H09 measurements of $\tilde{\psi}_k^p$ respectively. The 5th, 25th, 50th, 95th percentiles are shown as labelled, along with the maximum values of the periodograms in each spectral bin (labelled "max"). The vertical dashed lines indicate the $k = 0$ and $k = 10$ spectral bins, with frequencies given by $(k + 1)(5\text{ years})^{-1}$ Hz.

measured from each of the Case R12 and Case H09 realisations, along with the expected PSDs of δ_i^p and W_i. Arbitrarily chosen single realisations of $\tilde{\psi}_k^p$ in each case are also shown. The means of the periodograms in both cases are clearly the same, and equivalent to the predicted PSD, $S(f)$, given in Equation 5.18. Though this is as expected, it is both a check of the simulations of D_i^p, and a demonstration of the ability of the PSD estimation method to measure steep red spectra without bias.

In contrast, the distributions of $\tilde{\psi}_k^p$ in the frequency bins where $S(f) \approx S_{\mathrm{g,\,fit}}(f)$ are different between the cases. The single realisations of $\tilde{\psi}_k^p$ in each case shown in the top panel of Figure 5.4 begin to hint at these differences. In most spectral bins, the Case R12 periodogram is below the Case H09 periodogram. That this is a genuine trend is confirmed in the bottom panel of Figure 5.4. Here, I depict various "percentile" periodograms of the distributions of $\tilde{\psi}_k^p$ in each of Case R12 and Case H09. The percentile periodograms may be interpreted as contours of equivalent percentiles of the periodogram distributions in different spectral bins. For example, the "50%" percentile periodogram links the 50th percentile points of the distributions of periodogram values in each spectral bin. Below the 95th percentile, all Case R12 percentile periodograms lie below Case H09 percentile periodograms. This implies that in most spectral bins, most measurements of a periodogram in Case R12 will, like the individual ones shown in the top panel of Figure 5.4, be below most Case H09 periodograms. However, the 95th percentile periodograms are essentially equivalent, and the maximum value Case R12 periodogram is well above the maximum value Case H09 periodogram. These implied "tails" at high values in the Case R12 periodogram distributions in each spectral bin are highlighted in Figure 5.5, which depicts the distributions of the $\tilde{\psi}_k^p$ in the spectral bins indicated by the vertical lines in the bottom panel of Figure 5.4. The distributions are shown as the fractions of Case R12 and Case H09 periodograms at or above a given value.

Figure 5.5 also shows that the Case R12 periodogram distribution in the $k = 10$ spectral bin has a longer tail relative to the Case H09 distribution, as compared to the $k = 0$ spectral bin. This effect is also evident in the bottom panel of Figure 5.4, in that the fractional differences between the percentile periodograms are greater at the upper end of the GW-dominated frequency regime. This is consistent with the number of GW sources per spectral bin included in the Case R12 simulations going down with increasing frequency, as shown in Figure 5.2.

In summary, approximating $\tilde{\psi}_k^p$ with $\tilde{S}_k^p$ as discussed above, I find that:

- In most spectral bins, most realisations of $\tilde{S}_k^p$ in Case R12 will be below most realisations of $\tilde{S}_k^p$ in Case H09.
- The maximum possible values of $\tilde{S}_k^p$ in Case R12 will be higher than the maximum possible values of $\tilde{S}_k^p$ in Case H09.

5.5 Correlations between GW-induced ToA variation time series

Hellings & Downs (1983) showed that the *average* values of correlations between δ_i^p for different pulsars, for a stochastic, isotropic GWB, will always be given by the Hellings & Downs curve (Equation 3.66). The distributions of individual estimates of these correlations, much like the distributions of the PSD estimator $\tilde{S}_k^p$ considered above, will however depend on the nature of the GW signal. Here, I characterise the distributions of estimates of these correlations for multiple pulsar pairs in each case discussed above. I simulated 100 realisations of each of Case H09 and Case R12 ToAs as described in §5.3.2 for pulsars at the positions of each of the 20 pulsars timed by the PPTA, using the timing models specific to each pulsar (Manchester et al. 2013). For each realisation, the same set of GW sources was used to simulate GW-induced ToA variations for each pulsar. The pulsar distances were set arbitrarily between 0.1 kpc and 20 kpc.

I estimated the correlations between time-series δ_i^p and δ_i^q, ρ_{pq}, for each pulsar pair pq in each realisation of Case R12 and Case H09 ToAs. No autocorrelations were estimated. A frequency-domain estimation technique, based on the method of Yardley et al. (2011), was used. This technique will be detailed elsewhere (Hobbs et al., in preparation). I refer to the estimates of ρ_{pq} as $\tilde{\rho}_{pq}$. These estimates were performed using D_i^p time-series, rather than δ_i^p time-series, and the estimation technique was optimised using the expected PSD of D_i^p. The technique removes best-fit linear and quadratic terms from each D_i^p time-series using the standard TEMPO2 least-squares fitting algorithm. This mimics the effect of fitting pulse frequency and frequency-derivative terms to the simulated ToAs and then analysing the timing residuals.

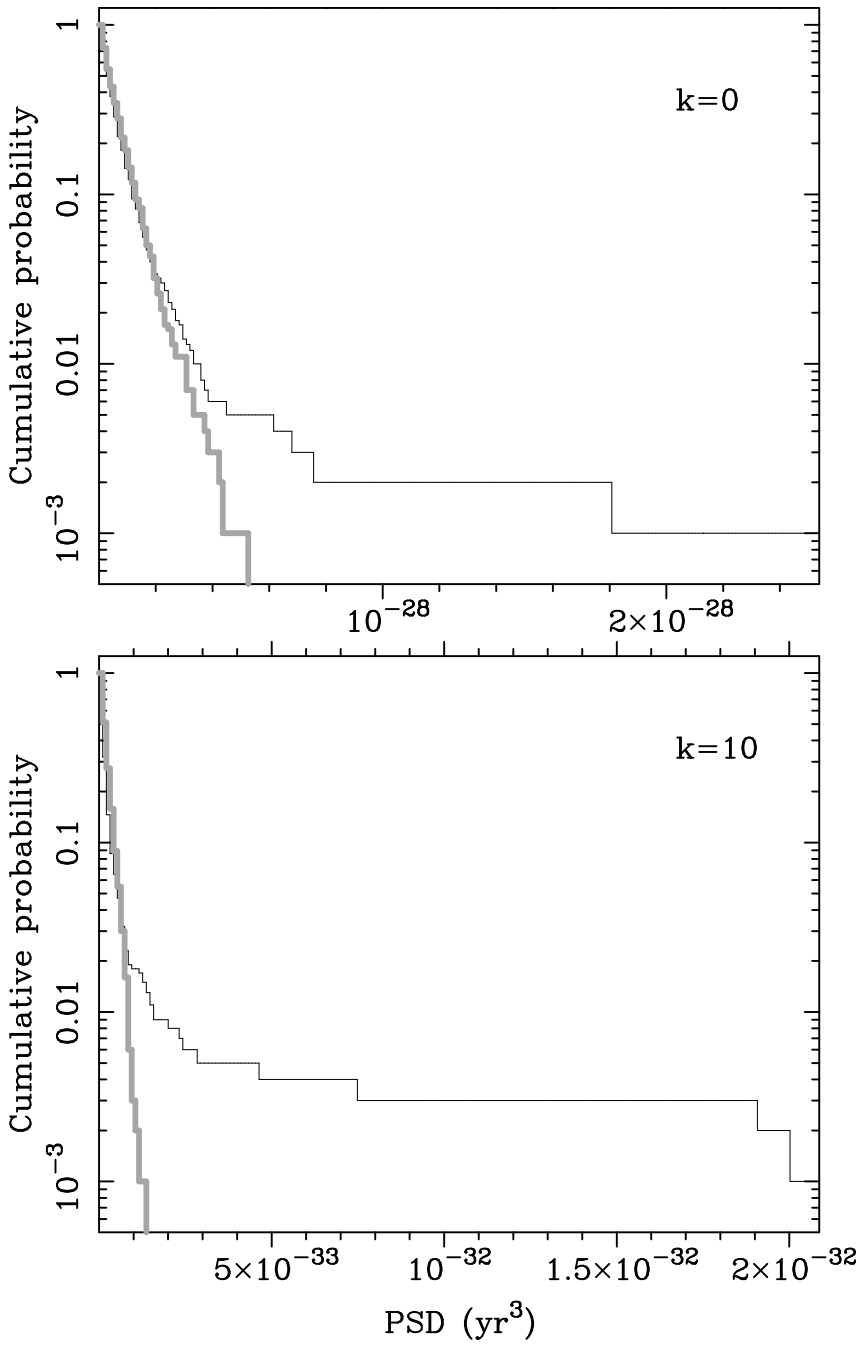

Figure 5.5: The distributions of 1000 measurements of $\tilde{\psi}_k^p$ in Case R12 (thin black lines) and in Case H09 (thick grey lines) in the $k = 0$ (top) and $k = 10$ (bottom) spectral bins. The distributions are shown as the fractions of realisations at or above a given value. The domains of both plots indicate the maxima and minima of the distributions.

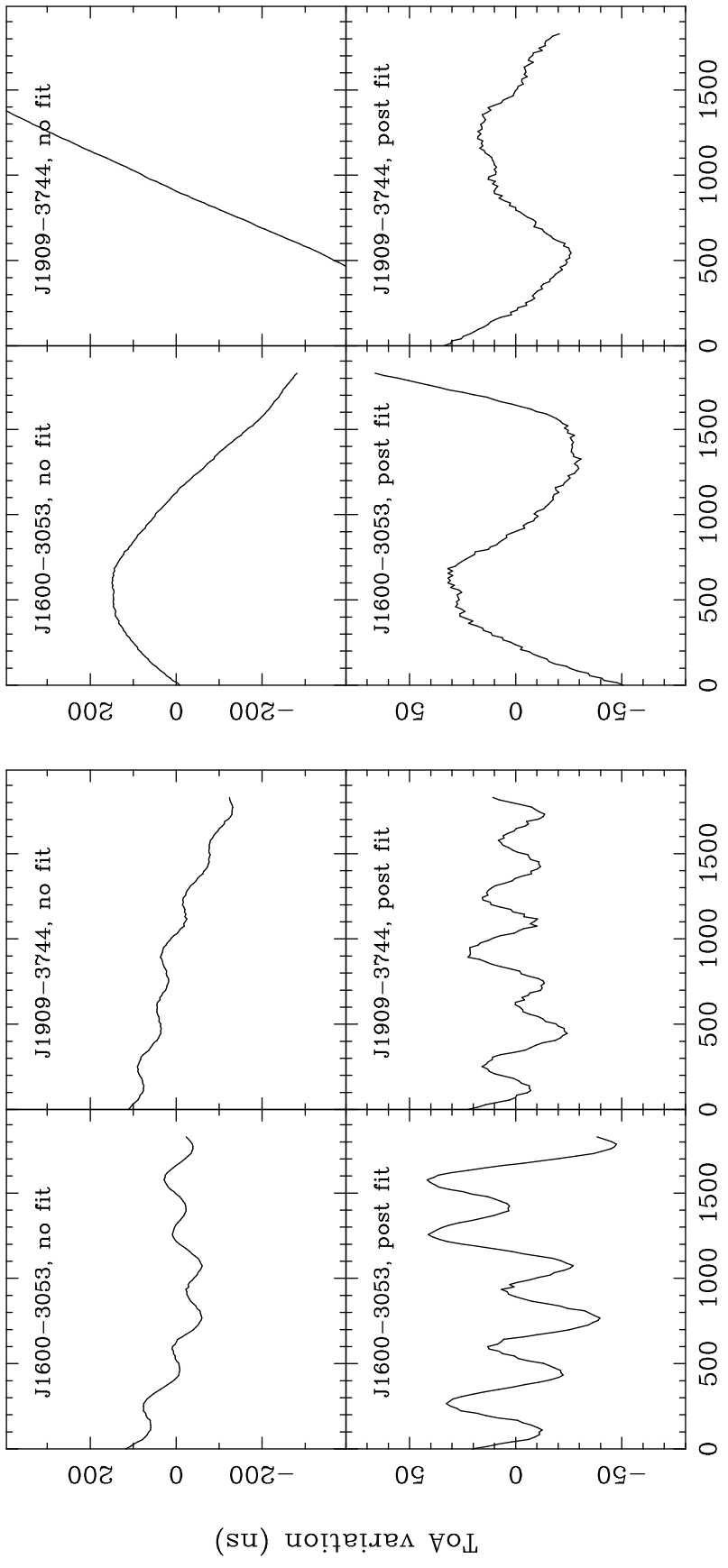

Figure 5.6: Two examples of simulated realisations of D_i^p for two pulsars: PSR J1600−3053 and PSR J1909−3744 (see text for details). The realisation in the left panel is affected by a strong individual GW source, whereas the realisations in the right panel is not. The lower plots show the D_i^p time-series from the corresponding upper plots with linear and quadratic terms removed.

Following the removal of linear and quadratic terms from one of the 100 Case R12 realisations of D_i^p, a single GW source was found to dominate the residual time-series. I show the corresponding D_i^p time-series for two of the 20 simulated pulsars for this realisation in Figure 5.6. The left panel of this Figure shows the large sinusoidal oscillations induced by the source, and the right panel shows example Case R12 realisations of D_i^p that are not dominated by an individual source. It is possible that an individual GW source with a period greater than the 5-year data span could dominate the realisations of D_i^p in the right-hand panel of Figure 5.6. The ToA variations induced by such a source would however be absorbed in the removal of the linear and quadratic terms from the D_i^p time-series.

I averaged all measurements of $\tilde{\rho}_{pq}$ for each pulsar pair pq from the Case R12 realisations, besides the one clearly dominated by an individual source. The realisation dominated by an individual source added a large amount of scatter to the average Case R12 correlations, and was left out of the average to enable a better (albeit conservative) comparison between the cases. I also averaged the Case H09 measurements of $\tilde{\rho}_{pq}$ for each pair pq in 99 arbitrarily-chosen realisations. The average measurements of $\tilde{\rho}_{pq}$ are shown for both cases in Figure 5.7. The functional form of the Hellings & Downs curve is recovered in both Case R12 and Case H09. However, the Case R12 estimates are significantly more scattered about the expected values of the correlations than the Case H09 estimates.

The increased scatter in the Case R12 correlations with respect to the Case H09 correlations in Figure 5.7 is caused by outlying estimates in only a few realisations of ToAs. This is shown in Figure 5.8, where I display the histograms of the $\tilde{\rho}_{pq}$ measurements between simulated ToA datasets for the positions of PSR J0437−4715 and PSR J0613−0200 in each case. Correlation estimates $|\tilde{\rho}_{pq}| > 1$ were possible because I normalised the estimated covariances between D_i^p time-series using the expected cross-PSD between the time-series. While most measurements in both cases are concentrated about the expected value of ρ_{pq}, a few Case R12 measurements are significantly displaced. This is consistent with the results of §5.4. I also stress that the large scatter of the estimator common to both cases is expected, and intrinsic to the GW signal.

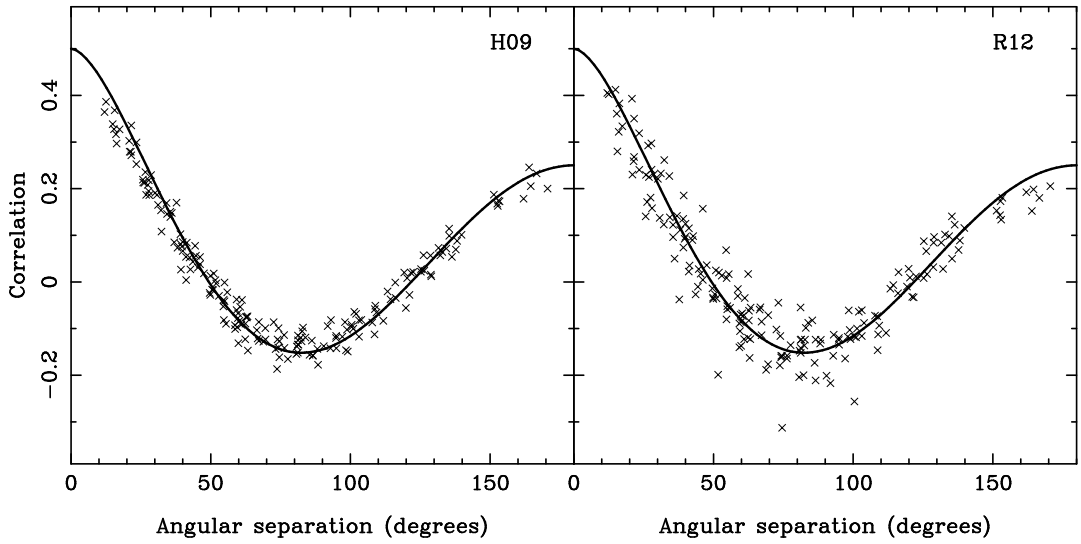

Figure 5.7: The estimated correlations between GW-induced ToA variations for simulated pulsars at the positions of the 20 Parkes PTA pulsars in Case H09 (left) and in Case R12 (right), plotted against the angular separations on the sky between each pair of pulsars. Each point represents an average over 99 realisations; in Case R12, one realisation including an extremely strong individual source was not included in the average. Linear and quadratic terms were removed from each ToA variation time-series. The solid curve is the expected Hellings & Downs curve given in Equation 4. As no autocorrelations were present, the maximum value of the Hellings & Downs curve is 0.5.

5.6 Discussion

I have shown that the ToA variations induced by GWs from the predicted binary SMBH population are not consistent with the model described in §5.2. The random Gaussian model for δ_i^p described in §5.2 and implemented by Case H09 is reasonable given that a large number of GW sources are expected to contribute to the GW-induced ToA variations. That is, the values of δ_i^p at all times t_i are the sums of many random variables. An argument based on the classical central limit theorem would suggest that δ_i^p would then be Gaussian random at every time t_i. It is apparent, however, that such a central limit theorem-based argument does not apply to the Case R12 realisations of δ_i^p. This is because of the nature of the GW sources contributing to δ_i^p in Case R12.

In Case R12, a few sources contribute most of the PSD of δ_i^p at every frequency, as shown in Figure 5.2. These sources are rare, because they are found at the high-h_0 tail of the $\Phi_{\rm fit}$ source distribution. The estimators I consider in this work, $\tilde{S}_k^p$ and $\tilde{\rho}_{pq}$, are dominated in Case R12 by a few GW sources that need not occur in every realisation of the δ_i^p time-series. This is why the distributions of these estimators are

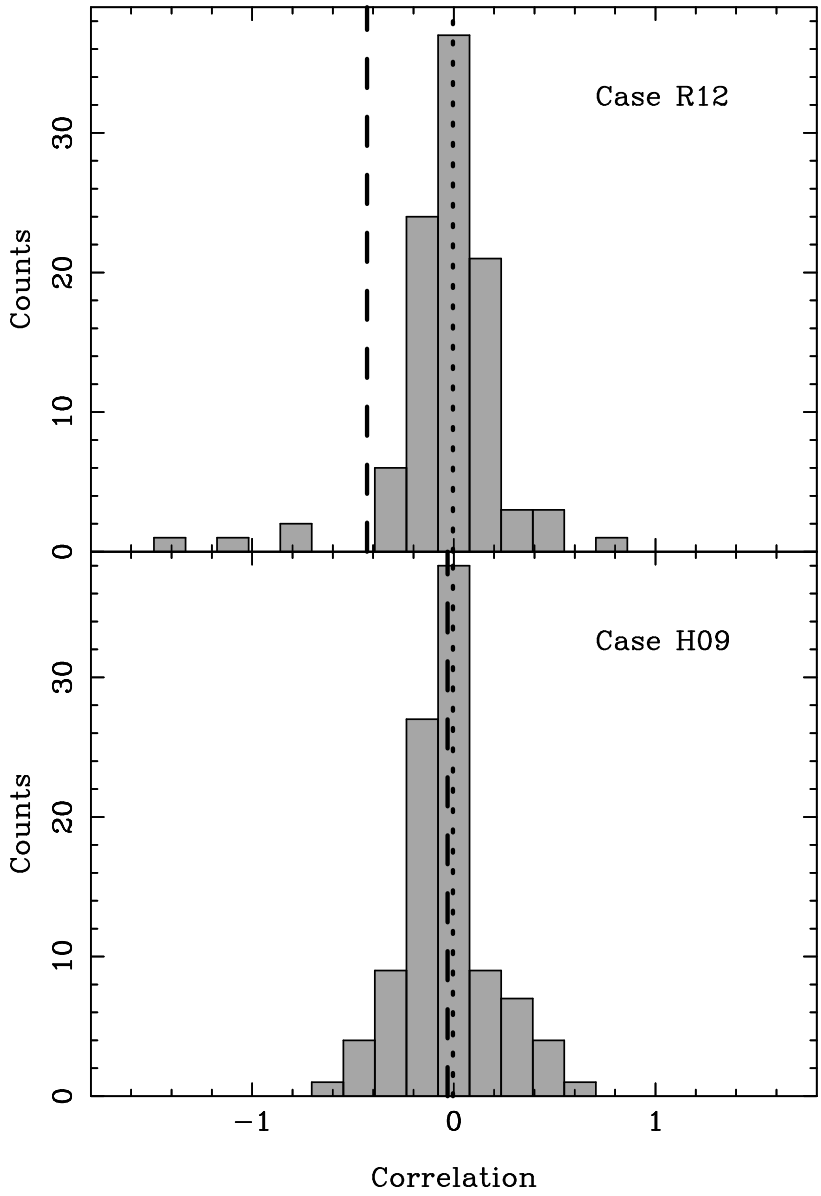

Figure 5.8: Histograms showing the distributions of measurements of the correlation estimator $\tilde{\rho}_{pq}$ for 100 simulated ToA datasets for PSR J0437−4715 and PSR J0613−0200, in Case R12 (*top*) and Case H09 (*bottom*). See the text for details of the simulations. The Case R12 realization that included an extremely strong single GW source, as discussed in the text, resulted in a measurement of $\tilde{\rho}_{pq} = -39.74$; this measurement is not shown in the Case R12 histogram. The vertical dashed line in each panel indicates the mean values of all 100 estimated correlations in each case, and the vertical dotted line indicates the expected value of the correlation, ρ_{pq}, for an angular separation of $\theta_{pq} = 49.8°$.

different between the cases. The quantities that I estimate, $S_g(f)$ and ρ_{pq}, are used to define the covariance matrix of the GW-induced ToA variations. I have therefore shown that the ToA variations induced by GWs from binary SMBHs are dominated by the effects of a few strong, rare sources and cannot be accurately modelled using the random Gaussian process discussed in §5.2.

The differences between the R12 and H09 models for the GWB from binary SMBHs are visually demonstrated in Figure 5.9. In the Figure, I show single realisations of the population of GW sources between frequencies of 10^{-9} Hz and 10^{-6} Hz simulated as described above for each model, and Aitoff-projected onto a sky map. I have added the values of h_s^2 for each source lying within the positional bounds of each pixel; the colour scheme represents the square roots of these sums. The brightness distributions in the two images are clearly different. Specifically, a few relatively bright individual sources are present in the R12 case, which are not present in the H09 case.

5.6.1 Implications of the results for experiments focused on a GWB

Current PTA data analysis techniques use assumptions about the statistics of δ_i^p to attempt to estimate or constrain the amplitude of the characteristic strain spectrum of GWs from binary SMBHs. I consider the implications of the results for a selection of techniques in turn. I assume, in this discussion, that the results for the statistics of GW-induced ToA variations would apply even if the normalisation of the GW characteristic strain spectrum $h_{\mathrm{c,\,fit}}(f)$, which I refer to as the GW amplitude,[4] were scaled up or down. Such a scaling could occur, for example, under different scenarios for whether coalescing SMBHs accrete gas before or after coalescence (Sesana et al. 2008b).

I summarise a few key techniques here:

- Jenet et al. (2005) describe a statistic which measures the degree of correlation between estimates of ρ_{pq} from ToA data, and the expected functional form of ρ_{pq}. The expected detection significance, which is estimated under the assumption that the GW-induced ToA variations are Gaussian random, satu-

[4] I make a distinction between this amplitude and A_{yr}, because $h_{\mathrm{c,\,fit}}(f)$ does not have exactly the same form as given in Equation 5.1.

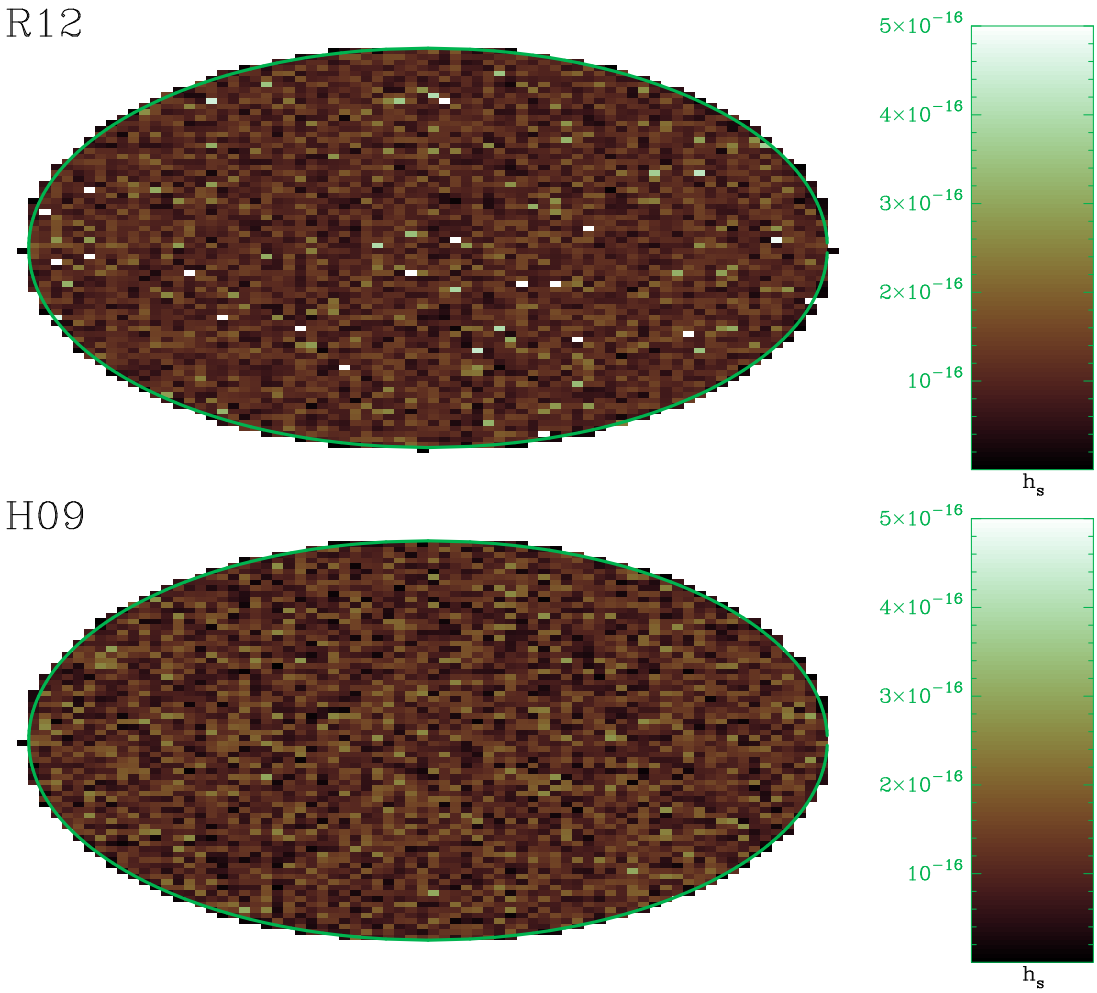

Figure 5.9: Realisations of the GW source populations in the R12 case (top) and in the H09 case (bottom), rendered onto an Aitoff projection of the sky. The individual source amplitudes, h_s, have been summed in quadrature within each pixel. The colour representations are shown to the right of each image; these are identical for both images.

rates at high values of the GW amplitude once the variance of the statistic is dominated by the stochasticity of the GW signal (see Figure 5.8 and related discussion).

- Jenet et al. (2006) constrain the amplitude of the GW characteristic strain spectrum from binary SMBHs by estimating the maximum possible GW signal present in measured data, under the assumption that the data could be modelled using a white noise process and GW-induced ToA variations. A statistic that estimates the GWB amplitude from individual pulsars was measured, and compared to the simulated distributions of the statistic for different GW amplitudes. The simulated statistic distributions were created from simulated ToA datasets with GW-induced ToA variations included using the TEMPO2 plugin GWbkgrd.

- van Haasteren et al. (2009) present a Bayesian parameter estimation method for the GW characteristic strain spectral amplitude.[5] van Haasteren et al. (2011) used this method to constrain the GW amplitude. This method requires an evaluation of the likelihood of the parameters used to model the ToA datasets, which include the GW amplitude. The likelihood is the probability distribution of the data given the model parameters. The GW amplitude is used to calculate the covariance matrix, $\mathbf{C_{pq}}$, between the GW-induced ToA variations for pulsars p and q (see Equations 7, 8). This covariance matrix in turn is used to define the PTA likelihood, assuming that the GW-induced ToA variations can be modelled as a random Gaussian process.

- Demorest et al. (2013) use a PTA likelihood similar to the work of van Haasteren et al. (2009) to constrain the GW amplitude by evaluating the distribution of a maximum likelihood estimator for the amplitude. They also use a method similar in concept to Jenet et al. (2005) to attempt to detect the GW signal from binary SMBHs.

First, the non-Gaussianity of the GW-induced ToA variations means that the estimate of the intrinsic GW-induced variance of the Jenet et al. (2005) statistic

[5]Though their method also estimates the spectral index of the GW characteristic strain spectrum, I assume marginalization over this parameter in the discussion here.

will be incorrect. This will affect estimates of the detection significance, particularly in the "strong signal regime", where the effects of GWs in the ToAs are large compared to all other noise processes. Second, the limit on the GW amplitude placed by Jenet et al. (2006) will be biased. The Jenet et al. (2006) limit was placed by finding the GW amplitude for which 95% of simulated statistic values were above the measured value. The distribution of their statistic derived using the simulations would be different. Ruling out a GW amplitude using the Jenet et al. (2006) technique does not necessarily rule out a GW signal corresponding to the simulations with the same confidence. Finally, the results indicate that the likelihoods evaluated by van Haasteren et al. (2009) and Demorest et al. (2013) will also be biased, leading to a similar effect on GW amplitude constraints made using their methods. A definitive statement on the magnitude of the consequences of the simulations for current constraints on the GWB amplitude cannot be made, however, without fully considering the various PTA data analysis methods. I demonstrate the consequences of my work for the GWB constraint of Shannon et al. (2013) in Appendix B to this Chapter.

5.6.2 Single GW source detection prospects

I have established that, given the specific model for the binary SMBH population that I consider, a few strong GW sources dominate the PSD of δ_i^p in every frequency bin of 5-year datasets. This means that the expected GW signal from binary SMBHs does not purely form an isotropic background. This result is analogous to the case of the extragalactic background light (Domínguez et al. 2011), where the summed electromagnetic radiation from AGN and star-forming galaxies is dominated by strong individual sources, behind which myriad further objects combine to form an apparently isotropic background too uniform to be resolved by current telescopes. I briefly consider the prospects for there being single sources of GWs that are detectable by PTAs. Various methods of detecting and characterising individual continuous sources of GWs with PTAs have recently been presented (Yardley et al. 2010; Boyle & Pen 2012; Corbin & Cornish 2010; Lee et al. 2011; Babak & Sesana 2012; Ellis et al. 2012; Arzoumanian et al. 2014; Zhu et al. 2014). There are, however, few predictions for the expected numbers of detectable sources. Sesana et al. (2009) analysed similar binary SMBH population models to those considered here to sug-

gest that a 5-year ToA dataset would include 5−10 single GW sources above the mean "stochastic background" level, mainly at GW frequencies greater than 10^{-8} Hz. Their definition of a *resolvable* source as one which has a (mean) strain amplitude that is greater than the mean background level is conservative. This is because a PTA is capable of spatial, as well as frequency resolution. The background contribution per spatial resolution element of a PTA will be less than the all-sky background level, resulting in a higher source amplitude to background ratio for a bright source located in the resolution element.

The exact number of resolvable GW sources given a GWB level for PTAs depends on the particular search method. For example, Boyle & Pen (2012) suggest that a PTA composed of N pulsars could resolve up to $2N/7$ sources per frequency bin. In Figure 5.10, I present a simple indication of the expected amplitudes of strong individual sources between frequencies of 3.2 nHz and 79 nHz (spectral bins with $0 \leq k \leq 10$) of the fiducial 5-year dataset. Using 300 Case R12 realisations of the GW source population, I found the mean strain amplitudes in each spectral bin of the strongest three GW sources. I express these amplitudes as multiples of the mean summed strain amplitude, $h_{\rm rest}$, of the remaining sources. The errors in the $h_{\rm rest}$ values were not included in the error bars as they were very small.

If I consider the sources besides the strongest three in a spectral bin to form a "background",[6] it is clear that, for spectral bins with $k > 1$ (frequencies greater than 16 nHz), three sources, on average, produce the same total strain amplitude as the remaining sources. Even for the first spectral bin, three sources are expected to produce more than half the total strain amplitude of the remaining sources. Indeed, the strongest source in the first spectral bin has an average strain amplitude that is $\sim 0.35 h_{\rm rest}$, which implies that a PTA which can resolve out two-thirds of the sky will detect equal contributions from the source and from the background.

Blind searches for single GW sources with PTAs are therefore likely to be important. PTA data analysis methods that attempt to detect an isotropic component will not optimally recover the entirety of the GW signal from binary SMBHs, and could perhaps miss a large component of the signal for some realisations of the GW source population. A careful consideration of the efficacy of GWB detection meth-

[6]This is by no means a rigorous definition of a background relative to the number of sources. The exact definition is dependent on the single source search method and the characteristics of the PTA.

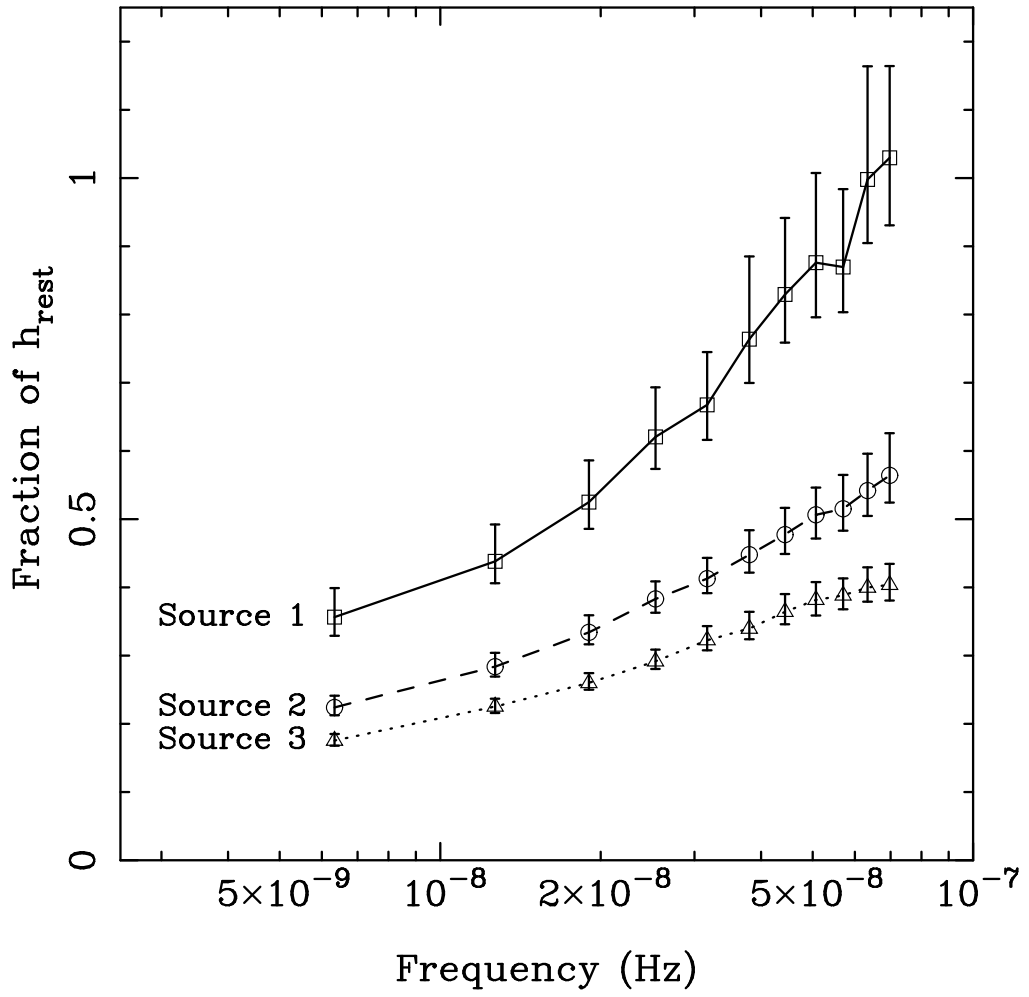

Figure 5.10: The average strain amplitudes of the three highest-amplitude binary SMBHs between frequencies of 3.2 nHz and 70 nHz (spectral bins with $0 \leq k \leq 10$) for each realisation of the population. The strain amplitudes are expressed as fractions of the mean summed amplitude of the remaining sources. I also show the 5th and 95th percentiles of the strain amplitudes, with their deviations from the means scaled down by a factor of 10. I made 300 realisations of the source population to produce this Figure. As indicated in the Figure, squares (the solid line) depict the mean amplitudes of the strongest sources, circles (the dashed line) depict the mean amplitudes of the second strongest sources, and triangles (the dotted line) depict the mean amplitudes of the third strongest sources.

ods as compared to search methods for single sources, given the predicted source characteristics, is required.[7]

5.6.3 Limitations of my modelling approach

My method of simulating $\delta_g(t)$ is limited by the fact that I do not account for uncertainties inherent in the galaxy evolution model itself. Sesana et al. (2008b) attempted an analysis of the uncertainties in the prediction of the GW signal from binary SMBHs by considering various scenarios for accretion before and after SMBH-SMBH coalescences, and by also considering uncertainties in the SMBH mass-galaxy scaling relations used to tune model parameters. In order to produce more realistic realisations of the expected population of binary SMBHs, I would need to randomise over such scenarios. This could possibly be accomplished by drawing the parameters of $\Phi_{\rm fit}$ from derived distributions in each realisation.

Another shortcoming of the approach towards modelling δ_i^p, and indeed of all predictions for the GW signal from binary SMBHs prior to the commencement of this thesis, is the assumption of circular orbits for all binaries. Recent work (e.g., Sesana 2010; Preto et al. 2011; Khan et al. 2011) suggests that binary SMBHs emitting GWs in the PTA frequency regime will have highly eccentric orbits. The candidate binary SMBH OJ 287 (e.g., Valtonen et al. 2008) is in fact modelled with an orbital eccentricity of $\sim$0.7. The GW waveform of an eccentric binary radiating in the PTA band spans many frequencies, and does not follow the frequency-time relation of Equation 3.33. Therefore, if most binary SMBHs radiating in the PTA band are eccentric, the predicted mean spectral slope of the characteristic strain spectrum will change. The binary eccentricity distribution would also need to be accounted for in the predictions of the statistics of GW-induced ToA variations. Also requiring further investigation are the effects of gas and stars on binaries.

5.7 Conclusions

I have used a sophisticated model for galaxy evolution (Guo et al. 2011) to predict the distribution of binary SMBHs radiating GWs in the PTA frequency band.

[7] Indeed, following the publication of this work in Ravi et al. (2012), a selection of other studies considered exactly this problem (e.g., Mingarelli et al. 2013; Taylor & Gair 2013; Cornish & Sesana 2013).

By drawing lists of GW sources from this distribution, I simulated the effects of GWs from binary SMBHs on 5-year pulsar ToA datasets. I compared these simulations (Case R12) with simulated pulsar datasets containing the effects of an equivalent-amplitude GW signal modelled as a random Gaussian process (Case H09). I estimated the PSDs of the simulated GW-induced ToA variation time-series, and the correlations between these time-series for different pulsars. I found that the distributions of the PSD estimators of the realisations of the GW-induced ToA variations are different between the cases in every frequency bin, although the mean estimated PSDs are the same in each case. While in Case R12 the estimated PSDs were concentrated at lower values than in Case H09, the Case R12 estimations extended to higher PSD values than the Case H09 estimations. I also found that the functional form of the Hellings & Downs curve is recovered on average in both cases. The correlations between the GW-induced ToA variation time-series for different pulsars in Case R12 were, however, significantly more scattered about the expected values than in Case H09. I interpret the results in terms of the influence of strong individual GW sources on the ToAs in Case R12.

I conclude the following:

1. The effects of GWs from binary SMBHs on pulsar ToAs cannot be accurately modelled using existing methods, i.e, as a random Gaussian process. This is because a few GW sources dominate the PSD of the GW-induced ToA variations at all frequencies, with reducing numbers of sources contributing equivalent PSD fractions in higher frequency bins. That is, *the GWB from binary SMBHs may be mildly anisotropic.*

2. The results directly affect existing PTA data analysis methods aimed at detecting or estimating the parameters of the GW signal from binary SMBHs. Projected detection significances will be biased.

3. The prospects for single GW source detection are strong. Individual sources could potentially be resolved in all GW-dominated frequency bins of a 5-year dataset.

Future searches for GW signals from binary SMBHs in pulsar datasets may hence need to be sensitive to both individual sources as well as a GWB.

5.8 APPENDIX A: The expected GWB amplitude from the G11 model

In this section, I present a calculation of the mean GWB amplitude predicted by the G11 model. While the work in this Chapter has shown that thinking of the GWB as isotropic and inducing Gaussian ToA variations may not be accurate, the expected energy density in GWs from binary SMBHs can still be calculated. However, this calculation is somewhat more difficult than evaluating $h_{\mathrm{c,\,fit}}(f)$ in Equation 5.15. There are two major reasons for this. First, the G11 model treatment of SMBHs, based on the work of Kauffmann & Haehnelt (2000) and Croton et al. (2006) as described in §2.3 by Equations 2.29 and 2.30, was designed to reproduce the relationship between galaxy spheroids and SMBH masses observed in the local Universe. However, the $M_{\bullet} - M_{\mathrm{bul}}$ relation used to tune the f_{BH} and κ_{AGN} parameters in Equations 2.29 and 2.30 respectively was the one derived by Häring & Rix (2004). As stated in §2.3, this relationship has been significantly updated, most recently by Kormendy & Ho (2013), hereafter KH13 in this Chapter, such that the mean observed ratio $\Gamma_{\mathrm{obs}} = M_{\bullet}/M_{\mathrm{bul}}$ is now thought to be a factor of 1.8 greater than that inferred by Häring & Rix (2004). Second, running the G11 model on the halo merger trees from the Millennium simulations essentially provides a realisation of the binary SMBH population within the simulation volumes. The original Millennium simulation volume corresponds to the comoving volume sphere extending to $z \sim 0.1$. Hence, at larger redshifts, it is unlikely that the Millennium simulation includes the rarest objects in the Universe, such as, for example, the remnants of the quasars observed at $z > 6$. As has been demonstrated in this Chapter, the rarest binary SMBHs may contribute a significant fraction of the expected energy density in GWs.

5.8.1 Updating the SMBH masses in the G11 model

Updating the $z \sim 0$ $M_{\bullet} - M_{\mathrm{bul}}$ relation resulting from the G11 model require the alteration of f_{BH} and κ_{AGN}. However, the large covariance between the two parameters (Mutch et al. 2013) implies that the parameters can be tuned to the new $M_{\bullet} - M_{\mathrm{bul}}$ relation without affecting the self-consistency of the model. I characterise the updated SMBH and spheroid mass sample by the mean SMBH

to spheroid mass ratio, $\Gamma_{\rm obs}$. Where available, I use spheroid mass estimates derived while including dark matter contributions to the gravitational potential. This was done in order to obtain the most accurate value of $\Gamma_{\rm obs}$ possible. Some authors (e.g., McConnell & Ma 2013) calculate the $M_\bullet - M_{\rm bul}$ relation using only spheroid mass estimates made only with dark matter contributions excluded, in order to obtain a self-consistent relation, whereas others, such as KH13, use the best spheroid mass estimates available in each case. No significant differences exist between the McConnell & Ma (2013) and the KH13 $M_\bullet - M_{\rm bul}$ relations.

To account for the revised sample of SMBH and bulge masses, I scaled the masses of SMBHs in the G11 model by a factor F, which is equivalent to adjusting the parameter $f_{\rm BH}$. This equivalence is physically justified for two reasons. First, it is thought that the vast majority of the mass of SMBHs in the local Universe has been built up through accretion in quasar phases, i.e., the masses of the first generation of black holes are relatively small. In fact, no SMBH seeds are included in the G11 model. Instead, upon the first merger experienced by a pair of galaxies, an SMBH with a mass given by Equation 2.29 is assumed to be created in the merger remnant. Second, SMBHs are at most a hundredth of the total baryon masses of their host galaxies, indicating that the contribution of SMBHs to the baryon masses of their host galaxies, and hence the amount of gas accreted in mergers, is largely independent of the SMBH mass. Together, these facts imply that an SMBH at any redshift in the G11 model, having undergone any number of accretion and coalescence episodes with other SMBHs, will have a mass that increases linearly with $f_{\rm BH}$. This was confirmed by examining the SMBH mass functions output by the Croton et al. (2006) semi-analytic model for different values of $f_{\rm BH}$.

Given that (a) the sample used to measure $\Gamma_{\rm obs}$ comprises only 35 SMBH-galaxy pairs, (b) individual mass measurements show large uncertainty, and (c) the $M_\bullet - M_{\rm bul}$ relation shows large intrinsic scatter, the value of F has significant uncertainty. To account for this uncertainty when calculating the strength of the GWB, we need to estimate the posterior probability distribution of the factor F given $\Gamma_{\rm obs}$, i.e., $\rho(F|\Gamma_{\rm obs})$. This is straightforward to evaluate using the Bayes Theorem:

$$\rho(F|\Gamma_{\rm obs}) \propto \rho(\Gamma_{\rm obs}|F)\rho(F) \tag{5.19}$$

where $\rho(\Gamma_{\rm obs}|F)$ is the probability density of obtaining $\Gamma_{\rm obs}$ for different values of

F, also referred to as the likelihood of F given $\Gamma_{\rm obs}$. I adopt a uniform prior in F (i.e., a uniform $\rho(F)$) within a reasonable range in F, $(0.8 < F < 3.2)$. I used a Monte Carlo technique to evaluate the posterior distribution $\rho(F|\Gamma_{\rm obs})$. For fixed F, I generated 10^5 random values of $\Gamma = M_\bullet/M_{\rm bul}$ from the G11 model. Each value was calculated using random selections of 35 SMBH-bulge pairs with the same bulge mass distribution as the sample of KH13. I also generated 10^5 random values of $\Gamma_{\rm obs}$ using the observational errors. The posterior distribution was then found by estimating the probability density of the distribution of $\log(\Gamma/\Gamma_{\rm obs})$ values at zero. This process was repeated for many values of F in the range $0.8 < F < 3.2$. The maximum likelihood estimate of F is 1.9, with the 5th and 95th percentiles of the posterior distribution $\rho(F|\Gamma_{\rm obs})$ lying at $F = 1.46$ and $F = 2.46$ respectively. This is consistent with the updated ratio of the normalisation of the $M_\bullet - M_{\rm bul}$ relation found by KH13.

5.8.2 Predicting $\Omega_{\rm GW}(f)$

The fitted distribution of binary SMBH GW sources from the G11 model, $\Phi_{\rm fit}(h_0, f)$, defined in Equation 5.12, cannot be scaled to correspond to an arbitrary F. This is because the derivative $\frac{dt}{df}$ used in converting the coalescence rate of pairs of SMBHs to the number of circular binaries per unit emitted GW frequency is dependent on the binary chirp masses and redshifts. These quantities are combined in defining the h_0-only distribution $\Phi_{\rm fit}$, and information about their individual values are lost. As I wish to convert the posterior predictive distribution for F, $\rho(F|\Gamma_{\rm obs})$, into a probability distribution of $\Omega_{\rm GW}(f)$, I need to calculate $\Phi_{\rm fit}(h_0, f)$ for numerous values of F. Rather than recalculating $\Phi_{\rm fit}(h_0, f)$ for each value of F by scaling the masses of the binary SMBHs from the G11 model, a significantly quicker method is to introduce a redshift-dependence to the fitted distribution of binary SMBHs.

For each interval between redshift snapshots in the Millennium simulation, with midpoint $Z_i = (z_i + z_{i+1})/2$ where the redshifts z_i are defined according to Equation 2.17, I counted the numbers, $n_i(h_0)$, of SMBH-SMBH coalescences in h_0 bins of width Δh_0. The distribution Φ is then approximated by

$$\Phi(h_0, z_i, f) \approx \frac{n_i(h_0)}{V_{\rm Mil}\Delta h_0} 4\pi \frac{d^2V_c}{d\Omega dz_i} \left[\frac{dz}{dt}\frac{dt}{df}\right]_{z=z_i} \tag{5.20}$$

where $V_{\rm Mil}$ is the Millennium simulation volume,and $4\pi\frac{d^2V_c}{d\Omega dz_i}$ is the comoving volume shell between redshifts z_i and z_{i+1}. I did not include coalescences from the Millennium-II simulation as they did not significantly affect the final result. Now, $\Omega_{\rm GW}(f)$ is given by

$$\Omega_{\rm GW}(f) = \frac{2\pi^2}{3H_0^2}\sum_i \int_{h_{\rm min}(z_i)}^{h_{\rm max}(z_i)} dh_0 \Phi(h_0, z_i, f) f^{13/3}. \tag{5.21}$$

It is important to model the counts $n_i(h_0)$ for each i in order to appropriately set the integration limits, $h_{\rm min}(z_i)$ and $h_{\rm max}(z_i)$. I summed each $n_i(h_0)$ over 1000 realisations of the binary SMBH coalescence lists derived from the G11 model, generated as described in §5.3.1, to form count distributions $n_{1000,i}(h_0)$. I then modelled these distributions by fitting broken power laws to $\log(n_{1000,i}(h_0))$ for each i using a least-squares fitting routine. The broken power law has the following functional form:

$$f_{\rm bpl}(h_0) = p_1 \left(\frac{h_0}{p_3}\right)^{p_2} \left(1 + \frac{h_0}{p_3}\right)^{p_4}, \tag{5.22}$$

with free parameters p_1, p_2, p_3 and p_4. However, given that the data are discrete counts, rather than data that are expected to be normally distributed about a particular regression curve, these fits will be biased. This bias will be particularly pronounced at the high-h_0 ends of the distributions. I therefore re-fit the distributions above the fitted p_3 values (i.e., above the turnover points of the broken power laws) assuming Poisson errors in every bin, as the function

$$f_{\rm pl}(h_0) = A \left(\frac{h_0}{p_3}\right)^{\alpha}. \tag{5.23}$$

I constrained A using the relation $f_{\rm pl}(p_3) = f_{\rm bpl}(p_3)$. I adopted the broken power-law fits for $h_0 < p_3$, and the simple power-law Poisson fits for $h_0 \geq p_3$. For some values of i, broken power-law fits were not possible, and for a few more cases, Poisson fits were not possible. The forms of the fitted functions are displayed in Figure 5.11.

In order to use the fits to the $n_i(h_0)$ distributions to calculate $\Omega_{\rm GW}(f)$, I needed to define the upper- and lower-h_0 limits of integration for each i, $h_{\rm min}(z_i)$ and $h_{\rm max}(z_i)$. The lower-h_0 bounds were set to be the h_0 value of a $10^6 - 10^6\,M_\odot$ binary at each redshift Z_i. The upper bounds were a bit more complicated to set. For each Z_i, I first found the largest h_0-value present in the simulation from the summed distributions

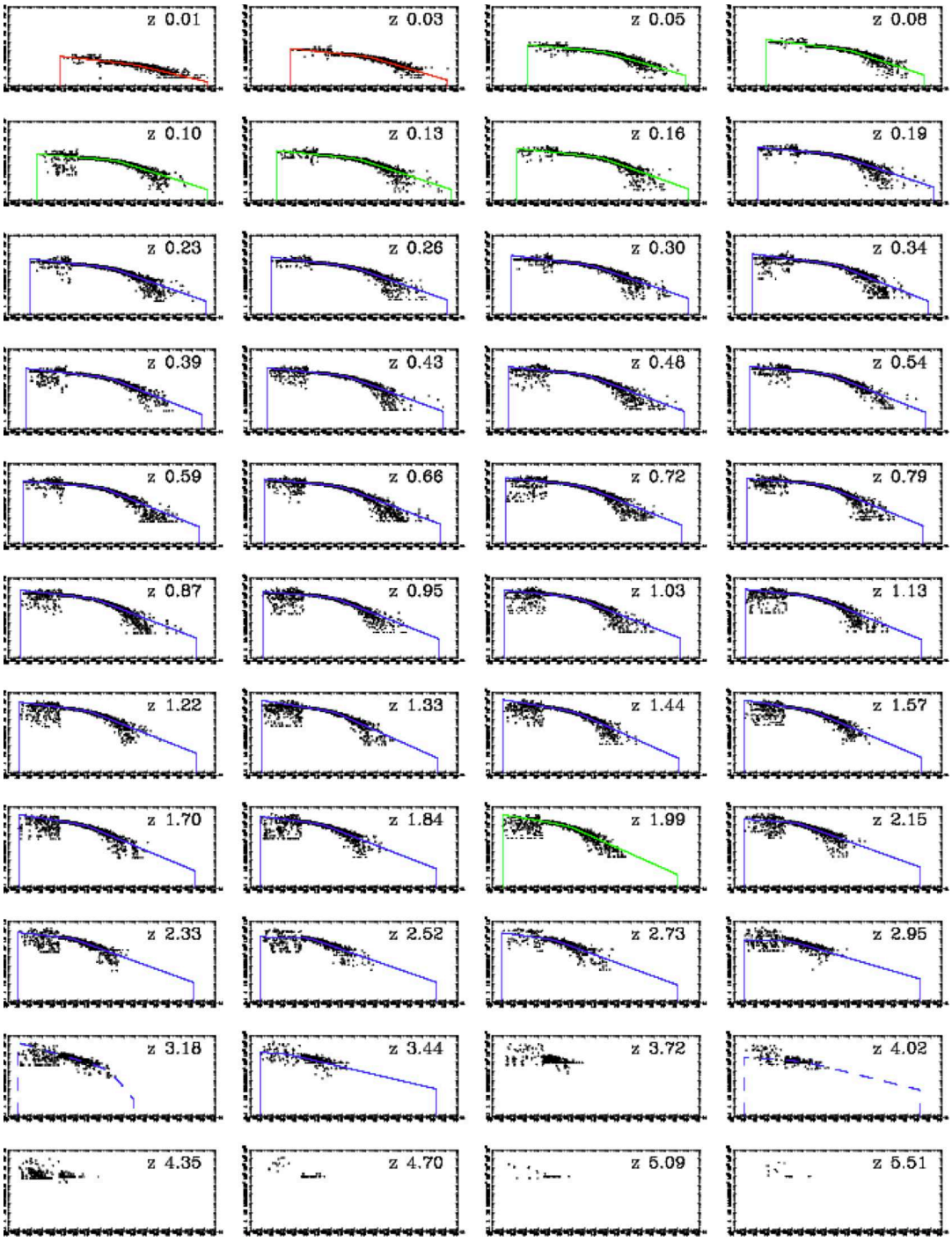

Figure 5.11: The forms of the distributions $n_{1000,i}(h_0)$ for each redshift Z_i for $i < 44$. The simulation results are shown as black dots, and, where present, the lines show the best-fit models. Red lines indicate upper-h_0 bounds on the models set by the simulations, green lines indicate upper bounds set by the h_0-value above which one source is expected, and blue lines indicate upper bounds set by the h_0-value of the most massive coalescence event at any redshift, scaled to the particular redshifts. Dashed lines indicate cases where Poisson fits were not possible, and the lack of any line in a panel indicates that no fit was done and that the simulation results themselves were used. In all other cases Poisson fits were applied to the models above the turning points of the broken power laws.

($n_{1000,i}(h_0)$), as well as the h_0-value above which one binary would be expected given the fitted functions. I then found the maximum of these two h_0-values, $h_{0,A}$. Then, I checked whether $h_{0,A}$ was less than the h_0-value of the most massive binary in the simulation at any redshift, scaled to the redshift of interest, $h_{0,B}$. If this was true, I adopted $h_{0,A}$ as the upper integration limit. Otherwise, I adopted $h_{0,B}$ as the upper integration limit. The most massive coalescence event in the model (i.e., in the 1000 realisations of the event list) has a chirp mass of $3.3 \times 10^{10}\, M_{\odot}$.

Note that I assumed that the h_0-limits of the integral in Equation 5.21 are independent of frequency, in contrast to the crude definition of $h_{0,\,\max}$ previously in this Chapter. I also neglected to remove binaries from the calculation when their GW emission frequencies corresponded to orbital separations within the innermost stable circular orbits. However, these effects are negligible for most GW frequencies of interest to PTAs (Wyithe & Loeb 2003a). For example, a binary system with two $10^{10} M_{\odot}$ SMBHs at a redshift of $z = 2$ will appear from the Earth to be radiating GWs at $f \sim 10^{-6}\,\mathrm{Hz}$ when it reaches the innermost stable circular orbit. Closer, less massive binary systems will be radiating at even higher frequencies; these frequencies are all above the standard PTA band. Hence, in the PTA band, I assume the standard power law form for the characteristic strain spectrum for all binary SMBHs (Equation 5.1), corresponding to $\Omega_{\mathrm{GW}}(f) \propto f^{2/3}$

I calculated $\Omega_{\mathrm{GW}}(f)$ in three cases using Equation 5.21. These computations are best compared by quoting the corresponding characteristic strain amplitudes at $f = f_{\mathrm{yr}}$, A_{yr}. First, by using the averaged distributions $n_{1000,i}(h_0)/1000$ with no fitting, I found $A_{\mathrm{yr}} = 6.63 \times 10^{-16}$. Next, when I used the broken power-law fits to the summed event distributions, and divide the fitted functions by 1000, neglecting Poisson fitting, and set the integration limits as described above, I found $A_{\mathrm{yr}} = 7.08 \times 10^{-16}$. Finally, when I used Poisson fitting for $h_0 > p_3$ at every redshift Z_i where the fits were possible, I found $A_{\mathrm{yr}} = 9.20 \times 10^{-16}$.

For a given SMBH mass scaling factor F, I shift each distribution $n_i(h_0)$ by setting $h_0 \rightarrow h_0 F^{-10/3}$. I also re-evaluate the upper limits of the h_0-integrations, leaving the lower limits fixed as above. For the maximum likelihood value $F = 1.9$, I find $A_{yr} = 1.57 \times 10^{-15}$. I found that the probability density function for $\Omega_{\mathrm{GW}}(f_{\mathrm{yr}})$,

given $\rho(F|\Gamma_{\rm obs})$, is well modelled by two log-normal distributions:

$$\begin{aligned}\rho_M(\log(\Omega_{\rm GW})) =& \frac{0.983}{\sqrt{2\pi(0.115)^2}}\exp\left(\frac{-(\log(\Omega_{\rm GW})+0.935)^2}{2(0.115)^2}\right) \\ &+ \frac{0.017}{\sqrt{2\pi(0.123)^2}}\exp\left(\frac{-(\log(\Omega_{\rm GW})+0.954)^2}{2(0.123)^2}\right).\end{aligned} \tag{5.24}$$

5.9 APPENDIX B: Constraining the amplitude of a non-Gaussian GWB

While the statistics of GWB-induced ToA variations are likely to be non-Gaussian, the PSD of these ToA variations, $S_g(f)$, is still described by Equation 3.65 and depends on the GWB amplitude $A_{\rm yr}$. This amplitude depends on the numbers of binary SMBHs with different masses and distances radiating GWs in the PTA frequency band, which are not known with any degree of certainty. The G11 semi-analytic galaxy formation model implemented in the Millennium simulation, coupled with recent measurements of SMBH and galaxy spheroid masses, results in a prediction of $A_{yr} = 1.57 \times 10^{-15}$, with approximately 0.1 dex (1σ) uncertainty. Different models, as listed in §3.1.4, result in different predicted values of $A_{\rm yr}$.

In §4.1, I described the best existing PTA upper limit on $A_{\rm yr}$ (Shannon et al. 2013), which was derived assuming that the GWB-induced ToA variations are a Gaussian random process. Here, using exactly the same techniques, I present a constraint on $A_{\rm yr}$ assuming that while the GWB amplitude is unknown, the statistics of the GWB-induced ToA variations are as described in this Chapter.

The estimator $\hat{A}^2$ derived by Shannon et al. (2013) (see Equation 4.3) is still optimal in the case of non-Gaussian GWB-induced ToA variations, because the PSD of the ToA variations remains the same as in the Gaussian case. A Fourier series description of the ToA variations also remains appropriate, as Iğ model these variations simply as the sum of sinusoidal signals contributed by numerous individual sources. In order to simulate distributions of the statistic $\hat{A}$ for different values of $A_{\rm yr}$, I replaced the use of the TEMPO2 plugin GWBkgrd by simply simulating the ToA variations contributed by lists of GW sources from the distribution $\Phi_{\rm fit}$, as described above in §5.3.2, but with the h_0-values of each source scaled by the factor $A_{\rm yr}/h_{\rm c,\,fit}(f_{\rm yr}$. Using an Anderson-Darling test, I found that distributions of $\hat{A}$ were not significantly different when sources were drawn from a h_0-domain contributing

50% of $S_g(f)$, rather than 90% of $S_g(f)$ as was described in §5.3.2. Using this technique, the resulting 95% confidence upper limit on $A_{\rm yr}$ was 2.7×10^{-15}.

As a test of the method of simulating non-Gaussian ToA variations, I also derived distributions of $\hat{A}$ where I generated samples of sources with h_0-values drawn from the fitted forms of $\Phi(h_0, z_i, f)$ (for z_i values where fits were possible), and frequencies drawn from a power-law distribution with slope $-11/3$. For values of $A_{\rm yr}$ corresponding to $F = 1$, 2, 3, I found no significant differences between the two techniques of generating distributions of $\hat{A}$, again using an Anderson-Darling test.

Chapter 6

Binary SMBH environments cause the GWB to be diminished

I assess the effects of supermassive black hole (SMBH) environments on the gravitational-wave (GW) signal from binary SMBHs. To date, searches with pulsar timing arrays for GWs from binary SMBHs, in the frequency band $\sim 1 - 100\,\mathrm{nHz}$, include the assumptions that all binaries are circular and evolve only through GW emission. However, dynamical studies have shown that the only way that binary SMBH orbits can decay to separations where GW emission dominates the evolution is through interactions with their environments. I augment an existing galaxy and SMBH formation and evolution model with calculations of binary SMBH evolution in stellar environments, accounting for non-zero binary eccentricities. I find that coupling between binaries and their environments causes the expected GW spectral energy distribution to be reduced with respect to the standard assumption of circular, GW-driven binaries, for frequencies up to $\sim 10\,\mathrm{nHz}$. Larger eccentricities at binary formation further reduce the signal in this regime. I also find that GW bursts from individual eccentric binary SMBHs are unlikely to be detectable with current pulsar timing arrays. The uncertainties in these predictions are large, owing to observational uncertainty in SMBH-galaxy scaling relations and the galaxy stellar mass function, uncertainty in the nature of binary-environment coupling, and uncertainty in the numbers of the most massive binary SMBHs. I conclude, however, that low-frequency GWs

from binary SMBHs may be more difficult to detect with pulsar timing arrays than previously thought. This Chapter describes my original work, which has largely been published (Ravi et al. 2014).

6.1 Introduction

A strong case exists for the formation of a binary SMBH upon the merger of a pair of galaxies hosting central SMBHs (Begelman et al. 1980). The central SMBHs sink in the merger remnant potential well through the action of dynamical friction, and form a bound binary when the mass within the orbit of the lighter SMBH is dominated by the heavier SMBH. As stars within the binary orbit are quickly ejected, the binary will decay further only if another mechanism to extract binding energy and angular momentum exists. Proposed mechanisms include slingshot scattering of stars on radial, low angular momentum orbits intersecting the binary (Frank & Rees 1976; Quinlan 1996; Yu 2002), and friction against a spherical Bondi gas accretion flow (Escala et al. 2004) or a circum-nuclear gas disk (e.g., Roedig et al. 2011). If the orbital decay process can drive the binary to a small separation, GW emission will eventually cause the binary to coalesce (e.g., Peters & Mathews 1963; Baker et al. 2006a).

The so-called 'final parsec problem' (Milosavljević & Merritt 2003), which refers to the long-standing difficulty in establishing how binaries could decay past the stage where dynamical friction becomes inefficient, is likely to now be solved in a purely stellar-dynamical context. While friction against gas has indeed been shown to efficiently shrink binary orbits, most of the massive, nearby binary SMBH systems that are of relevance to this thesis are likely to exist in "dry" (gas-poor) mergers. Currently, however, it appears that numerical simulations of galaxy mergers with realistic stellar distributions result in sufficient stars on radial orbits to efficiently drive binary decay to the GW-dominated regime (e.g., Khan et al. 2012), although recent work (Vasiliev et al. 2014) has questioned the numerical convergence of these findings.

Merging dark matter halos follow parabolic trajectories (e.g., van den Bosch et al. 1999; Willott 2011), implying large initial eccentricities (typically ~ 0.6, Hashimoto et al. 2003) for the orbits of SMBHs sinking towards galaxy merger remnant centres. Steep stellar density gradients in merging

galaxies may reduce this eccentricity; indeed, some models suggest that binary SMBHs are likely to be close to circular upon formation (Casertano et al. 1987; Polnarev & Rees 1994; Hashimoto et al. 2003). Slingshot interactions between binaries and individual stars again grow the eccentricities (Quinlan 1996; Sesana et al. 2006; Berentzen et al. 2009; Khan et al. 2012), because binaries spend more time, and hence lose more energy, at larger separations. Roedig et al. (2011) found that binary SMBHs embedded in massive self-gravitating gas disks will have large eccentricities, between 0.6 and 0.8, at the onset of GW-dominated evolution. In the GW-dominated regime, binaries have long been predicted to quickly circularise (e.g., Peters & Mathews 1963), although more modern strong-field analyses find that the circularisation is somewhat slower than first suggested (Glampedakis et al. 2002; Gair & Glampedakis 2006).

There is no direct observational evidence for the existence of binary SMBHs. However, the GW emission from binaries prior to coalescence is an unambiguous signature of their existence. This emission is being searched for with PTAs. PTAs target both a stochastic, isotropic GWB from binary SMBHs (e.g., Yardley et al. 2011; van Haasteren et al. 2011; Demorest et al. 2013) and GWs from individual binary systems (e.g., Arzoumanian et al. 2014; Zhu et al. 2014). The summed GW signal from all binary SMBHs in the Universe is expected to approximate an isotropic background, although, as I demonstrated in Chapter 5, individual binaries are potentially detectable at all frequencies within the PTA band. However, most current predictions for the spectral shape, statistical nature and strength of the GWB from binary SMBHs assume that all binaries are in circular orbits, and losing energy and angular momentum only to GWs. These assumptions correspond to the well-known power-law GWB characteristic strain spectrum from binary SMBHs that is proportional to $f^{-2/3}$, where f is the GW frequency (see Equation 5.1).

Here, I present an examination of the properties of the GW signal from binary SMBHs given a realistic model for binary orbital evolution. I use a semi-analytic model for galaxy and SMBH formation and evolution (Guo et al. 2011, hereafter G11 in this Chapter) implemented in the Millennium simulation (Springel et al. 2005) to specify the SMBH-SMBH coalescence rate, and augment this with a framework (Sesana 2010) for the evolution of binary SMBHs in stellar environments. I neglect gas-driven binary evolution. This is because massive galaxy mergers at low redshifts, which are expected to dominate the total energy density in GWs from binary SMBHs

(e.g., Sesana et al. 2004), will typically be early-type and dry (e.g., Yu et al. 2011).

Two key phenomena in binary SMBH evolution affect the summed GW signal relative to the case of circular binaries evolving under GW emission alone[1]:

1. Interactions between binary SMBHs and their environments will accelerate orbital decay compared to purely GW-driven binaries, reducing the time each binary spends radiating GWs within a particular frequency interval. This may reduce the energy density in GWs at the lower end of the PTA frequency band.

2. While circular binaries emit GWs at the second harmonics of their orbital frequencies, eccentric binaries emit GWs at multiple harmonics (Peters & Mathews 1963). Given a population of binary SMBHs, this is expected to transfer GW energy density from lower frequencies in the PTA frequency band to higher frequencies (Enoki & Nagashima 2007).

I consider the effects of both these phenomena on the GW signal from binary SMBHs relative to the circular, GW-driven case. I also examine the possibility of detecting bursts of GWs from individual eccentric, massive binaries. In §6.2, I outline the binary population model. I describe predictions for the summed GW signal in §6.3, along with a discussion of uncertainties in the model. I consider the possibility of detectable GW bursts in §6.4. Finally, I summarise the results and implications for PTA studies in §6.5 and present the basic conclusions in §6.6. In this Chapter, I adopt a concordance cosmology consistent with the Millennium simulation (Springel et al. 2005), with $\Omega_M = 0.25$, $\Omega_b = 0.045$, $\Omega_\Lambda = 0.75$, and $H_0 = 73\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

6.2 Description of modelling methods

6.2.1 The binary SMBH population at formation

To begin, consider binary SMBHs with component masses $M_1 \geq M_2$, orbital semi-major axes a_0 and eccentricities e_0, embedded in isotropic, unbound cuspy stellar distributions with velocity dispersions σ. Quinlan (1996) found that binary hardening caused by slingshot interactions with individual stars becomes effective at

[1] I refer to this as the "circular, GW-driven case" throughout this Chapter.

binary component separations of a_h, defined in Equation 3.6. I assume that dynamical friction is effective in driving SMBHs in galaxy mergers to mean separations a_h (e.g., Callegari et al. 2009; Khan et al. 2012), and consider binaries at this stage to be newly formed.

I adopt a simple, one-parameter distribution for the eccentricities of binary SMBHs at formation, based on the postulate that the semi-major and semi-minor axes of the orbits (a_0 and b_0 respectively) are each log-normally distributed. This is justified because (a) while many stellar encounters influence the values of a_0 and b_0, the effects of these encounters on the parameter values are heterogeneous, and (b) both a_0 and b_0 are strictly positive.[2] In general, log-normal distributions are used to model positive-definite random variables that are influenced by many multiplicative effects of differing magnitudes (i.e., heterogeneous effects). The central limit theorem implies that the *product* of a large number of finite-variance positive random variables will approximately have a log-normal distribution.

I hence model the ratio b_0/a_0 using a probability density function given by

$$F_0\left(\frac{b_0}{a_0}, w_0\right) = \begin{cases} \sqrt{\frac{2}{\pi}}\frac{a_0}{b_0 w_0}\exp\left[-\left(\frac{\ln(\frac{b_0}{a_0})}{w_0\sqrt{2}}\right)^2\right], & \frac{b_0}{a_0} \leq 1 \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

Here, w_0 is the free parameter; larger values of w_0 correspond to typically larger binary eccentricities, and $w_0 = 0$ corresponds to a population of circular binaries. I do not consider any variation of w_0 with binary component masses or redshift, because there are no strong motivations for such variations. The eccentricity of a binary at a_h is given by $e_0 = \sqrt{1-(b_0/a_0)^2}$.

Let $\boldsymbol{\zeta_0} = [M_1, M_2, e_0]$ be a vector of parameters of binaries at formation. I denote the distribution of binaries in these parameters as $D_{\boldsymbol{\zeta_0}}[N(\boldsymbol{\zeta_0}, z)]$. In this notation, the multivariate density function for a parameter vector $\mathbf{X}$ with components X_i indexed by an integer i is given by

$$D_{\mathbf{X}}[N] \equiv \prod_i \frac{\partial[N]}{\partial X_i}.$$

Binaries at formation have semi-major axes $a_0 = a_h/(1+\sqrt{1-e_0^2})$.

[2]See Gaddum (1945) for a discussion of the ubiquity of log-normal distributions in nature.

I use the results of the semi-analytic model of G11 to specify $D_{\zeta_0}[N(\zeta_0, z)]$. As outlined in Chapter 4, the G11 results can be used to predict the coalescence rate of binary SMBHs. For this work, I only use coalescences with both M_1 and M_2 greater than $10^6 M_\odot$, and only draw from the implementation of the G11 model in the Millennium simulation. I scale all SMBH masses by a factor of 1.9 to account for recent SMBH and galaxy bulge measurements.

I count the coalescing pairs of SMBHs in bins of z, M_1 and M_2 (with widths Δz, ΔM_1, and ΔM_2 respectively) within the entire Millennium simulation box. Binaries are also randomly assigned values of e_0 using Equation 6.1. In this Chapter, I consider four different initial binary eccentricity distributions defined by $w_0 = 0$, 0.1, 0.35, 0.93. Denoting the binary counts for different values of z, M_1, M_2 and e_0 by the discrete distribution $n(\zeta_0, z)$,

$$\frac{d}{dz}[D_{\zeta_0}[N(\zeta_0, z)]] \approx \frac{n(z, \zeta_0)}{V_{\rm Mil}\Delta z \Delta M_1 \Delta M_2 \Delta e_0}, \tag{6.2}$$

where $V_{\rm Mil}$ is the comoving volume of the Millennium simulation box (Springel et al. 2005). I average the distribution $n(\zeta_0, z)$ over different specifications of the coalescence list from the G11 model, as described in Chapter 4, and 1000 realisations of the initial e_0-distribution. I do not fit an analytic function to $n(\zeta_0, z)$; I discuss the possible consequences of this below.

I relate the dark matter halo virial velocities, $V_{\rm vir}$, of galaxies in the G11 model to spheroid stellar velocity dispersions σ_c (Baes et al. 2003; Marulli et al. 2008). For each bin of z, M_1 and M_2, I find the average velocity dispersions of recently-merged galaxies in the G11 model hosting an SMBH of mass $M_1 + M_2$. I use these values to specify a_h for each bin of the discrete distribution $n(\zeta_0, z)$.

6.2.2 Evolution of binary SMBH orbits to the GW regime

I assume that all SMBH binary orbits decay through interactions with fixed, isotropic, unbound cuspy stellar backgrounds, and through GW emission. This scenario has been extensively studied numerically by Quinlan (1996), Sesana et al. (2006) and Sesana (2010), and is described in §3.1. I assume a power-law stellar density distribution within the binary gravitational influence radius for all galaxies prior to mergers (see Equations 3.1 and 3.3). For the majority of this

Chapter, I additionally assume a stellar density profile power-law index of $\gamma = 1.5$ corresponding to a mild stellar cusp. I consider variations in these assumptions further below.

I evolve the binary eccentricities, e, and semi-major axes, a, through scattering by unbound stars and loss of energy and angular momentum to GWs using expressions for $\frac{da}{dt_r}$ and $\frac{de}{dt_r}$ from Equations (8) and (9) of Sesana (2010). The effects of the ejection of stars that are *bound* to the SMBHs (Sesana et al. 2008a) are significant only for binary separations greater than a_h, and I hence neglect this phenomenon.

I use the fits of Sesana et al. (2006) for the rates of evolution of binary semi-major axes and eccentricities based on numerical scattering experiments (the 'H' and 'K' coefficients respectively from Tables 1 and 3 of Sesana et al. 2006, defined by Equations 3.8 and 3.10 respectively in Chapter 3). I log-interpolate the published values at binary component mass ratios of interest. As Sesana et al. (2006) only provide rates of semi-major axis evolution for circular binaries, I assume here that the rate of semi-major axis evolution at a given semi-major axis is independent of eccentricity. This approximation leads to the semi-major axis evolution rate being underestimated by at most 20% for the most eccentric binaries (see Figure 3 of Sesana et al. 2006). I also only use the seven values for the initial binary eccentricities (i.e., e_0) considered by Sesana et al. (2006); see their Table 3. These are 0, 0.15, 0.3, 0.45, 0.6, 0.75 and 0.9.

By numerically integrating the expressions for $\frac{da}{dt_r}$ and $\frac{de}{dt_r}$ for each combination of $\boldsymbol{\zeta_0}$ and z, I first calculate the binary eccentricities, $e_{\rm GW}$, at a rest-frame orbital frequency of 10^{-12} Hz. Binaries with this orbital frequency emit negligible GW power in the PTA frequency band. The orbital frequency of a binary is given by

$$f_{\rm orb} = \frac{1}{2\pi}\left(\frac{G(M_1+M_2)}{a^3}\right)^{1/2}. \tag{6.3}$$

Letting $\zeta_{\rm GW} = [M_1, M_2, e_{\rm GW}]$, I hence form the distribution function of binaries with orbital frequencies of 10^{-12} Hz, $D_{\zeta_{\rm GW}}[N(\zeta_{\rm GW}, z)]$, from the distribution of binaries at formation. If a binary at formation has $f_{\rm orb} > 10^{-12}$ Hz, I do not evolve the binary backwards in time to an orbital frequency of 10^{-12} Hz.

To then specify the population of GW-emitting binary SMBHs, I need to calculate the numbers of binaries with different orbital frequencies. The GW luminosity, L, per unit frequency, f_r, of a binary SMBH depends on the masses M_1 and M_2,

the eccentricity e, and the orbital frequency $f_{\rm orb}$ (Peters & Mathews 1963). The functional form of $L(f_r, \zeta)$ is given in Equation 3.36. I now define a new parameter vector $\zeta = [M_1, M_2, e, f_{\rm orb}]$.

The distribution $D_{\zeta_{\rm GW}}[N(\zeta_{\rm GW}, z)]$ can be used to specify the distribution function $D_\zeta[N(\zeta, z)]$ using a continuity equation similar to Equation 3.52:

$$\frac{d}{df_{\rm orb}}\left[\frac{df_{\rm orb}}{dt_r}D_\zeta[N(\zeta,z)]\right] = -\frac{d}{dt_r}[D_{\zeta_{\rm GW}}[N(\zeta_{\rm GW},z)]]\delta(f_{\rm orb}), \tag{6.4}$$

where $\frac{d}{dt_r}[D_{\zeta_{\rm o}}[N(\zeta_{\rm GW}, z)]]$ is the number of coalescences of binary SMBHs with parameters $\zeta_{\rm GW}$ per unit proper time t_r. The derivative $\frac{df_{\rm orb}}{dt_r}$ is equivalent to $\frac{df_{\rm orb}}{da}\frac{da}{dt_r}$, where $\frac{da}{dt_r}$ is given in Equation (15) of Sesana (2010). The solution is

$$D_\zeta[N(\zeta,z)] = -\frac{d}{dt_r}[D_{\zeta_{\rm GW}}[N(\zeta_{\rm GW},z)]]\left(\frac{df_{\rm orb}}{dt_r}\right)^{-1} \tag{6.5}$$

$$= -\frac{dz}{dt_r}\frac{d}{dz}[D_{\zeta_{\rm GW}}[N(\zeta_{\rm GW},z)]]\left(\frac{df_{\rm orb}}{dt_r}\right)^{-1} \tag{6.6}$$

I also associate each value of $f_{\rm orb}$ with a unique value of e by further integrating the expression for $\frac{de}{dt_r}$ from Sesana (2010).

Then, from Equation 3.49,

$$\Omega_{\rm GW}(f) = \int_0^\infty \left[\int ... \int_\zeta \frac{fL(f_r)D_\zeta[N(\zeta,z)]}{\rho_c c^2 H(z)(1+z)} dM_1 ... df_{\rm orb}\right] dz \tag{6.7}$$

Recall that $f = f_r/(1+z)$. For consistency with other works, I calculate the characteristic strain spectrum, defined as

$$h_c(f) = f^{-1}\left(\frac{3H_0^2}{2\pi^2}\Omega_{\rm GW}(f)\right)^{1/2}. \tag{6.8}$$

I perform the integral in Equation 6.7 over $f_{\rm orb}$ between $10^{-12} - 10^{-5}$ Hz. The upper orbital frequency limit corresponds to GW emission that is outside the PTA frequency band, even for binaries at high redshifts. For eccentric binaries, I consider radiation up to the 100th harmonic of $f_{\rm orb}$ (Peters & Mathews 1963). I assume that binaries reach their last stable orbits at separations of three Schwarzschild radii of the more massive SMBH (Hughes 2002), and neglect GW emission at smaller separations.

6.3 Predictions for the characteristic strain spectrum

6.3.1 Results

As stated above, I consider four different initial eccentricity distributions: $w_0 =$ 0, 0.1, 0.35, 0.93. Recall that the $w_0 = 0$ case corresponds to all binaries being circular. Initially circular binaries are not expected to become eccentric because of conservations of the Jacobian integral of motion (e.g., Sesana 2010). The probability mass functions of the binary eccentricities, e_0, at a_h in the three cases with $w_0 > 0$ are shown in Figure 6.1. For comparison, I also show in the bottom panel of Figure 6.1 a 'thermal' probability mass function for e_0, derived from the probability density function $f_{e_0} = 2e_0$ for $0 \leq e_0 \leq 1$. This would be expected if binary systems followed a purely Maxwell-Boltzmann distribution of energies (e.g., Ambartsumian 1937), as is roughly the case for galactic stellar binaries (Duquennoy & Mayor 1991).

In Figure 6.2, I plot the characteristic strain spectra for each initial eccentricity distribution. Also depicted is the prediction in the circular, GW-driven case (i.e., for $\frac{da}{dt_r}$ including only GW-driven orbital decay for all a). This latter prediction corresponds to the standard $h_c(f) \propto f^{-2/3}$ power-law. In order to help highlight the physical effects at work, Figure 6.3 shows the characteristic strain spectra for each assumed w_0 contributed by binaries with combined masses in the ranges $10^{6.5} M_\odot - 10^{10} M_\odot$ and $10^{10} M_\odot - 10^{11} M_\odot$ respectively.

The model I utilise for interactions between binaries and their stellar environments results in an attenuation of $h_c(f)$ in the PTA frequency band compared to the $f^{-2/3}$ power-law obtained in the circular, GW-driven case. For $w_0 = 0$, the signal is attenuated at frequencies $f \lesssim 10^{-8}\,\mathrm{Hz}$. At these frequencies, stellar interactions are the dominant binary orbital decay process, increasing $\frac{df_{\mathrm{orb}}}{dt_r}$ in Equation 6.6 and reducing the number of binaries observed per unit orbital frequency. For increasing w_0, the signal is further attenuated at low frequencies, although a slight ($\sim 0.01\,\mathrm{dex}$), increasing excess is present at frequencies between $10^{-8}\,\mathrm{Hz}$ and $10^{-7}\,\mathrm{Hz}$. This is caused by two effects: eccentric binaries evolve faster than circular binaries, and eccentric binaries radiate GWs at higher harmonics of their orbital frequencies than circular binaries.

The 'substructure', or two bumps, in the characteristic strain spectra is a direct

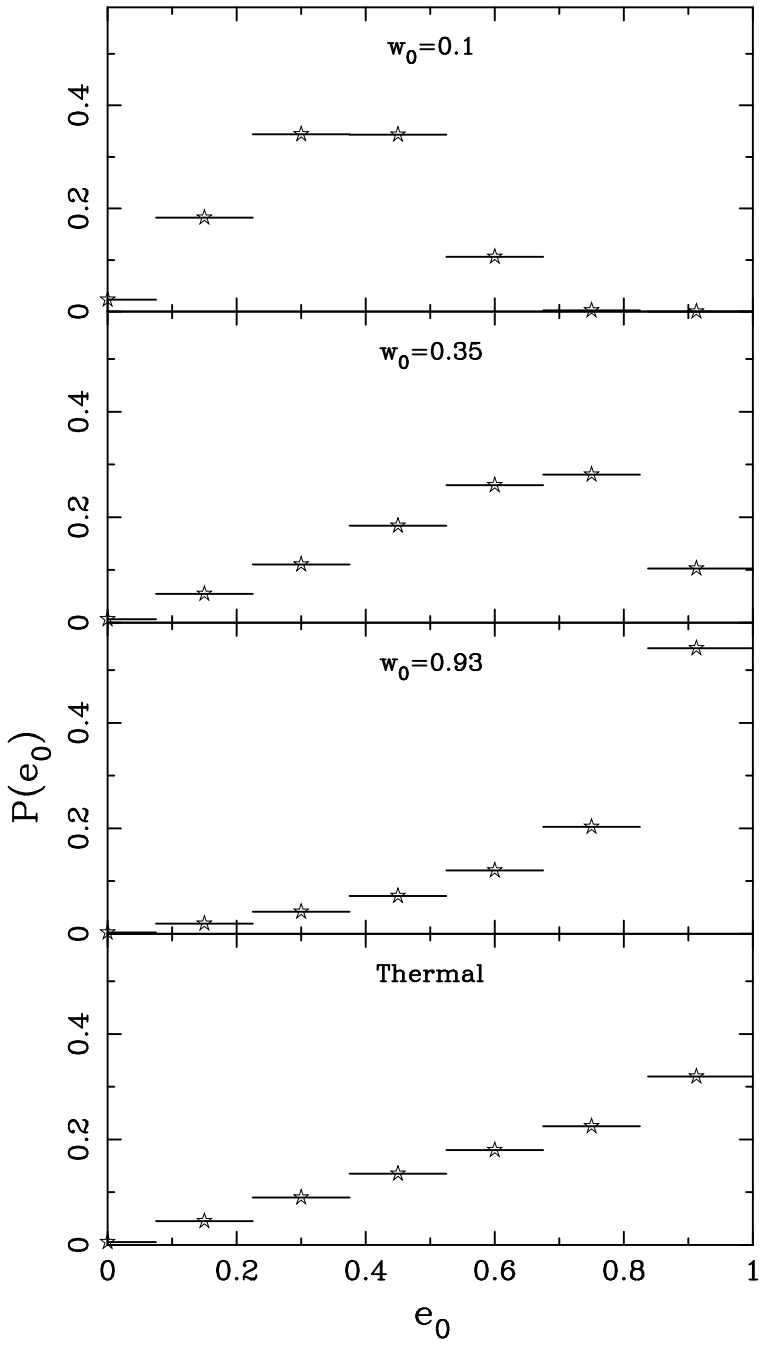

Figure 6.1: Probabilities, $P(e_0)$, of obtaining different values of e_0 (indicated by stars) for three initial eccentricity distributions defined by $w_0 = 0.1$, 0.35, 0.93 in Equation 6.1 (top three panels), and for a thermal eccentricity distribution (bottom panel). The values of e_0 correspond to those considered by Sesana et al. (2006); see text for details.

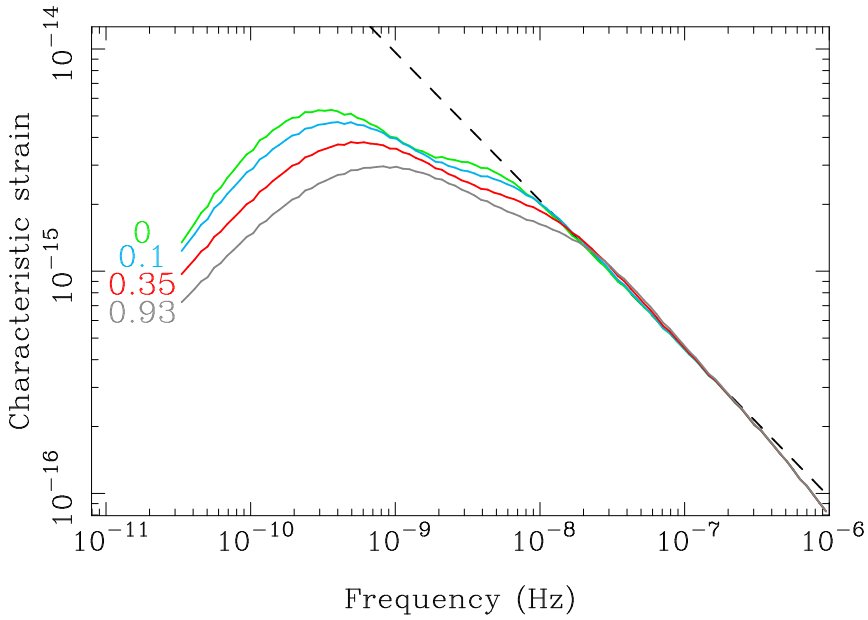

Figure 6.2: The solid lines depict characteristic strain spectra for $w_0 = 0$ (green), $w_0 = 0.1$ (blue), $w_0 = 0.35$ (red) and $w_0 = 0.93$ (grey); the w_0 values for each line are given at the left of the plot. All curves were calculated assuming a stellar density profile index of $\gamma = 1.5$. The black dashed line is the characteristic strain spectrum assuming circular orbits and purely GW-driven evolution for all SMBH binaries.

consequence of the mass-distribution of the binaries in the model. If $D_{\zeta_0}[N(\zeta_0, z)]$ were smooth and analytic, the characteristic strain spectra would have only one clear peak. Here, however, I evaluate this distribution from the G11 semi-analytic model outputs, which results in the distribution being incomplete at the high-mass end. These gaps in the distribution lead to the two apparent peaks in the characteristic strain spectra.

As is evident in Figure 6.3, the first peaks of the spectra in Figure 6.2 are dominated by the highest-mass binaries, whereas the second peaks are dominated by lower-mass binaries. This is because the evolution of the highest-mass binaries begins to be GW-driven at lower frequencies than for less massive binaries. There are expected to be very few binaries in the (combined) mass range $10^{10} M_\odot - 10^{11} M_\odot$; only ~ 50 with $f_{\rm orb} \geq 10^{-12}$ Hz are expected to be present in the observable Universe according to the G11 model. In contrast, $\sim 5 \times 10^6$ binaries are expected the the range $10^{6.5} M_\odot - 10^{10} M_\odot$. The effects of sparsity in $D_{\zeta_0}[N(\zeta_0, z)]$ are discussed further below.

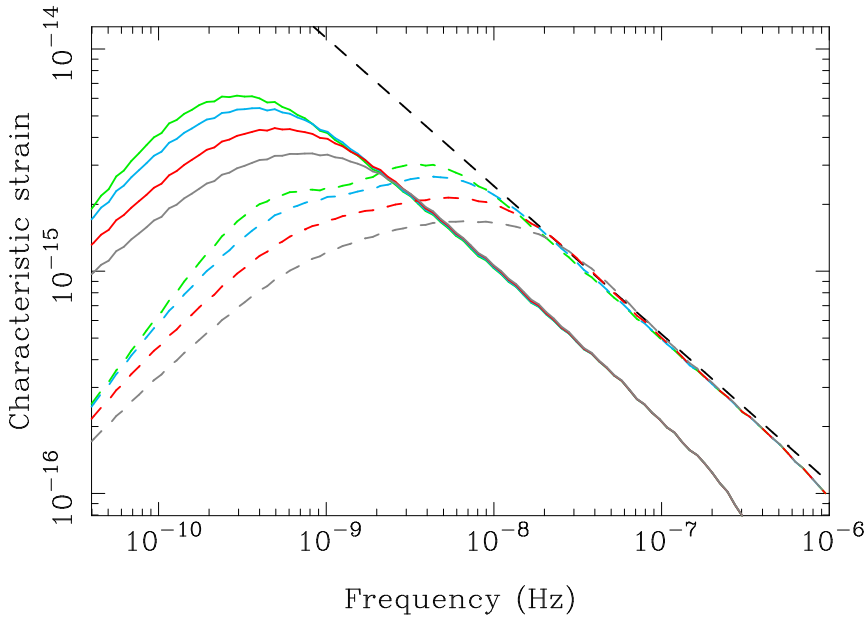

Figure 6.3: Characteristic strain spectra contributed by binaries with total masses in the range $10^{6.5}M_{\odot}-10^{10}M_{\odot}$ (dashed curves) and in the range $10^{10}M_{\odot}-10^{11}M_{\odot}$ (solid curves). The colours represent different values of w_0 as in Figure 6.2; note that the orders of the low- and high-mass curves from top to bottom correspond to increasing w_0 as in Figure 6.2. I again show the characteristic strain spectrum for all SMBH binaries assuming circular orbits and purely GW-driven evolution as a black dashed line.

6.3.2 Comparison with previous work

My results for the GW characteristic strain spectra from an eccentric binary SMBH population are broadly consistent with similar studies published elsewhere (Enoki & Nagashima 2007; Sesana 2013a). Both these works find spectra which depart from the standard power law of the circular, GW-driven case at frequencies $f < 10^{-8}\,\mathrm{Hz}$. The results of Sesana (2013a) for binaries with eccentricities at formation of 0.7 are in fact very similar to ours (see their Figure 2), with slight substructure evident along with the slight excess for $f > 10^{-8}\,\mathrm{Hz}$.

The results for $w_0 = 0$, however, differ somewhat from those of Sesana (2013a). Whereas the maximum separation between the zero-eccentricity and high-eccentricity curves (the red solid and dashed curves in Figure 2 of Sesana 2013a) is approximately 0.5 dex, the maximum difference between my curves for $w_0 = 0$ and $w_0 = 0.93$ in Figure 6.2 is 0.35 dex. I also find similarly-shaped spectra for all w_0, whereas Sesana (2013a) has a clear single peak in their zero-eccentricity curve.

The differences between my results and those of Sesana (2013a) for $w_0 = 0$ are caused by the nature of the respective binary SMBH mass distributions used. As dis-

cussed above, if the mass-distribution of binary SMBHs ($D_{\zeta_0}[N(\zeta_\mathbf{0}, z)]$) were smooth and analytic, which is the case in Sesana (2013a), only a single peak is expected. The reason for the similarity between my results and those of Sesana (2013a) for non-zero eccentricities may be because of some discreteness in the eccentric binary SMBH distribution used by Sesana (2013a), as evidenced by the jagged nature of their strain spectrum at low frequencies.

The characteristic strain spectrum I predict in the circular, GW-driven case is $\sim 0.15\,\mathrm{dex}$ lower than the results presented in §5.8. This difference is because I do not fit an analytic function to the discrete binary distribution $n(\zeta_\mathbf{0}, z)$.

6.3.3 Uncertainties in the model predictions

In this subsection, I describe the key uncertainties in the prediction of $h_c(f)$, which are summarised in Figure 6.4. I consider in turn the accuracy of the model predictions for SMBH demographics and coalescence rates and for the rate of evolution of binary systems, and the effects of incomplete high-mass binary SMBH distributions.

6.3.3.1 SMBH demographics and coalescence rates

The merger rate of massive galaxies predicted by galaxy formation models (Bertone et al. 2007) implemented in the Millennium simulation (Springel et al. 2005) has been shown to be consistent with observational estimates at redshifts $z < 2$ (Bertone & Conselice 2009). Marulli et al. (2008) found that the model matches the observed quasar bolometric luminosity function at redshifts $z \leq 1$ for a variety of assumed quasar lightcurves. This, together with the reproduction of the local SMBH-galaxy scaling relations, suggests that the rate of formation of massive binary SMBHs at low redshifts is satisfactorily reproduced by the G11 semi-analytic model, which is used as the basis for Chapters 5 and 6. Furthermore, I find that the characteristic strain spectrum expected in the $w_0 = 0$ case for binaries with combined masses $M_1 + M_2 > 2 \times 10^8 M_\odot$ at redshifts $z \leq 1$ has a maximum disparity with the unrestricted spectrum of 0.02 dex. Hence, I argue that model robustly predicts the contribution to the GW signal from massive, low-redshift binaries, which are likely to dominate the total GW signal (see also Wyithe & Loeb 2003a; Sesana et al. 2004).

However, there remain a range of theoretical uncertainties. For example, the G11 model treatment of SMBHs does not include physically-motivated prescriptions for SMBH formation (e.g., Haiman 2013), SMBH ejection caused by gravitational recoil following the coalescence of binary systems (e.g., Gerosa & Sesana 2014), and does not account for any mass accreted onto SMBHs in merging galaxies prior to coalescence (e.g., Van Wassenhove et al. 2012).

There are also specific observational uncertainties in tuning the semi-analytic model. The current sample of SMBH and host galaxy bulge mass measurements, which is used to tune the quasar-mode SMBH accretion efficiency, allows for a 1σ confidence interval of ~ 0.2 dex in the SMBH masses. Similarly, the galaxy stellar mass function predicted by the G11 model is matched to Sloan Digital Sky Survey observations in the nearby Universe (e.g., Li & White 2009). These observations have a $\sim$0.2 dex systematic uncertainty, with negligible contribution from cosmic variance (Li & White 2009), which corresponds (to first order) to a ~ 0.3 dex uncertainty in the galaxy merger rate.

The uncertainty in SMBH masses corresponds to a ~ 0.3 dex uncertainty in $\Omega_{\rm GW}(f)$, while the uncertainty in the merger rate translates directly to the range of predictions for $\Omega_{\rm GW}(f)$ allowed by the observed galaxy stellar mass function. Combining both ranges results in a 0.4 dex (1σ) uncertainty in $\Omega_{\rm GW}(f)$, which corresponds to a 0.2 dex uncertainty in $h_c(f)$.

6.3.3.2 The binary evolution model

In this Chapter, I assume that all galaxies hosting SMBHs have spherically-symmetric central stellar density profiles that are power-law functions of radius, r, following Sesana (2010). As detailed in Equation 3.1, the stellar density, $\rho(r)$, is proportional to $r^{-\gamma}$ within the binary influence radius, where I have hitherto assumed $\gamma = 1.5$. These profiles are equivalent to the central (asymptotic) behaviour of the Dehnen (1993) stellar potential and density models, which correspond well to high-resolution observations of the centres of nearby galaxy bulges (Faber et al. 1997). The assumption of a universal γ is, however, not in agreement with observations, which typically show $1 \lesssim \gamma \lesssim 2$, with $\gamma = 1$ corresponding to the most extreme 'core' galaxies and $\gamma = 2$ corresponding to the most extreme 'power-law' galaxies (Dehnen 1993; Faber et al. 1997). Furthermore, 'core' galaxies are

generally more massive, early-type systems with more massive SMBHs, and 'power-law' galaxies are generally less massive, late-type systems with less massive SMBHs (e.g., Faber et al. 1997; Kormendy & Ho 2013). While I do not attempt to correlate γ with galaxy properties from the G11 model, I show in Figure 6.4 characteristic strain spectra in the $w_0 = 0$ case for $\gamma = 1$ and $\gamma = 2$. The logarithmic differences between the spectra for these γ-values and the $w_0 = 0$ spectrum for $\gamma = 1.5$ may be applied only approximately to the spectra for other w_0-values, because varying γ varies both the rate of semi-major axis decay and the rate of eccentricity evolution for binaries.

The model that I use (Sesana et al. 2006; Sesana 2010) for binaries evolving through separations less than a_h due to interactions with fixed stellar backgrounds is qualitatively similar to the results of recent numerical simulations of dry (i.e., free of dynamically significant gas) galaxy merger remnants (Khan et al. 2012). However, as I show in Appendix A of this Chapter (§6.8), it is apparent that the model I use includes stronger stellar-driven orbital decay than the simulations of Khan et al. (2012). This is despite the assumption that the rate of semi-major axis evolution is independent of binary eccentricity. This is not surprising, because the assumption of a fixed stellar background is qualitatively equivalent to the assumption of a full stellar loss-cone (cf. Quinlan & Hernquist 1997; Sesana 2010). Hence, the model I use maximises binary orbital decay rates, in particular for spherically symmetric stellar distributions.

Hence, the model likely overestimates the effects of stellar interactions on the binary SMBH population. The numerical simulations of Khan et al. (2012) suggest that the orbital frequencies at which binary SMBH evolution begins to be predominantly GW-driven are up to 0.45 dex less than the corresponding frequencies that result from the model I use. This implies that the frequency below which the characteristic strain spectrum turns over from the $h_c(f) \propto f^{-2/3}$ power law may be up to 0.45 dex lower than I predict.

While the $V_{\rm vir} - \sigma$ relation that I assume is established in the local Universe (Baes et al. 2003), it has not been studied at higher redshifts. Given the expected decrease in the stellar mass in a halo of a given mass with increasing redshift (Moster et al. 2010), it is possible that I am overestimating the velocity dispersions of the stellar cores of merger remnants beyond the local Universe. This would imply that higher-redshift binaries decay more slowly than I predict, again increasing the

low-frequency parts of the presented characteristic strain spectra. Further work is required to quantify the magnitude of this increase.

Finally, while the assumed functional form of the e_0-distribution (Equation 6.1) is physically motivated, there may be some correlation between the orbital eccentricities of binaries with separations a_h, and their masses and redshifts. Additionally, a variety of studies find physical reasons for binaries to be quite circular upon formation (Casertano et al. 1987; Polnarev & Rees 1994; Hashimoto et al. 2003), which suggests that low-w_0 values may be preferred. Current numerical simulations (e.g., Khan et al. 2012) have not been run with a sufficient range of initial conditions to provide conclusive results on this point.

6.3.3.3 Accounting for discreteness in the binary SMBH distributions from the G11 model

Given the distribution $n(\boldsymbol{\zeta_0}, z)$ of binary SMBHs, I have examined the *expected value* of the GW characteristic strain spectrum. However, the binary SMBH distribution that I use is not exactly the distribution expected from the G11 semi-analytic model implemented within the Millennium simulation. The Millennium simulation provides a single realisation of the dark matter halo merger history within a large comoving volume, and the G11 prescriptions specify properties of the galaxies, and SMBHs, associated with the halos. To form the discrete binary distribution $n(\boldsymbol{\zeta_0}, z)$, I count the numbers of binary SMBHs forming in the entire Millennium volume between redshift snapshots in bins of M_1 and M_2, assigning values of e_0 to each binary according to my e_0-distribution. However, despite the large volume, $n(\boldsymbol{\zeta_0}, z)$ is poorly populated for high M_1 and M_2 at every redshift. To estimate the *expected* nature of this distribution, statistical modelling is required. This was done in §5.8, where I considered the circular, GW-driven case, and found that the modelling resulted in a characteristic strain spectrum increased by 0.15 dex. However, the distribution $n(\boldsymbol{\zeta_0}, z)$ has two more dimensions (an extra mass dimension and the eccentricity dimension) than that considered in the previous Chapter, which significantly complicates the modelling. Instead, I simply consider it possible that the strain spectrum I have derived may be up to 0.15 dex larger.

A qualitatively similar effect was pointed out by (Sesana et al. 2008b), who compared characteristic strain spectra generated from realisations of the binary SMBH

population of the Universe to the spectrum expected on average, in the circular, GW-driven case. Whereas the average spectrum was a power law proportional to $f^{-2/3}$, individual realisations had a lower amplitude at higher frequencies. This was because the numbers of binaries radiating GWs at a given frequency (per unit frequency) decreases with increasing frequency, implying that, for example, there is a frequency above which the expected number of sources is less than unity. However, the correct model for the *average* characteristic strain spectrum still had $h_c(f) \propto f^{-2/3}$ for all f, despite all realisations of the spectrum being below this power-law at high frequencies. This situation is analogous to my suggestion of an increase in the characteristic strain spectrum if the average behaviour of $n(\boldsymbol{\zeta_0}, z)$ were correctly modelled.

I also do not attempt here to describe the statistical nature of the GW signal, as was done in the previous Chapter in the circular, GW-driven case. I showed that the GW signal that was mildly non-Gaussian, with individual sources dominant at all GW frequencies of interest to PTAs. Realisations of $\Omega_{\rm GW}(f)$ at a particular frequency f would have a larger variance in the non-Gaussian case than in the Gaussian case. As discussed in §6.3.1, environment-driven binary evolution causes the highest-mass binaries to dominate $\Omega_{\rm GW}(f)$ at low frequencies to a greater extent than in the purely GW-driven case. This causes the low-frequency substructure in the characteristic strain spectra for all w_0 in Figure 6.2. My results, however, suggest a more general conclusion: that, at low frequencies, environment-driven binary evolution causes the variance in realisations of $\Omega_{\rm GW}(f)$ to be significantly increased relative to the assumption of only GW-driven evolution. Including this increased variance in $\Omega_{\rm GW}(f)$ at low frequencies in the calculation of PTA upper bounds on $\Omega_{\rm GW}(f)$ would cause these constraints to be further degraded relative to constraints based on the work in Chapter 5.

6.3.3.4 Synthesis of uncertainties in $h_c(f)$

I refer the reader to Figure 6.4, where I show an approximate 1σ confidence interval on the characteristic strain spectrum according to the model I describe. This interval represents the uncertainty in the expected value of the signal, not the realisation-to-realisation uncertainty. The interval encompasses the maximum possible ranges of w_0 and γ, and also includes observational uncertainties in the SMBH-bulge mass relation and in the galaxy stellar mass function (see §5.3.3.1).

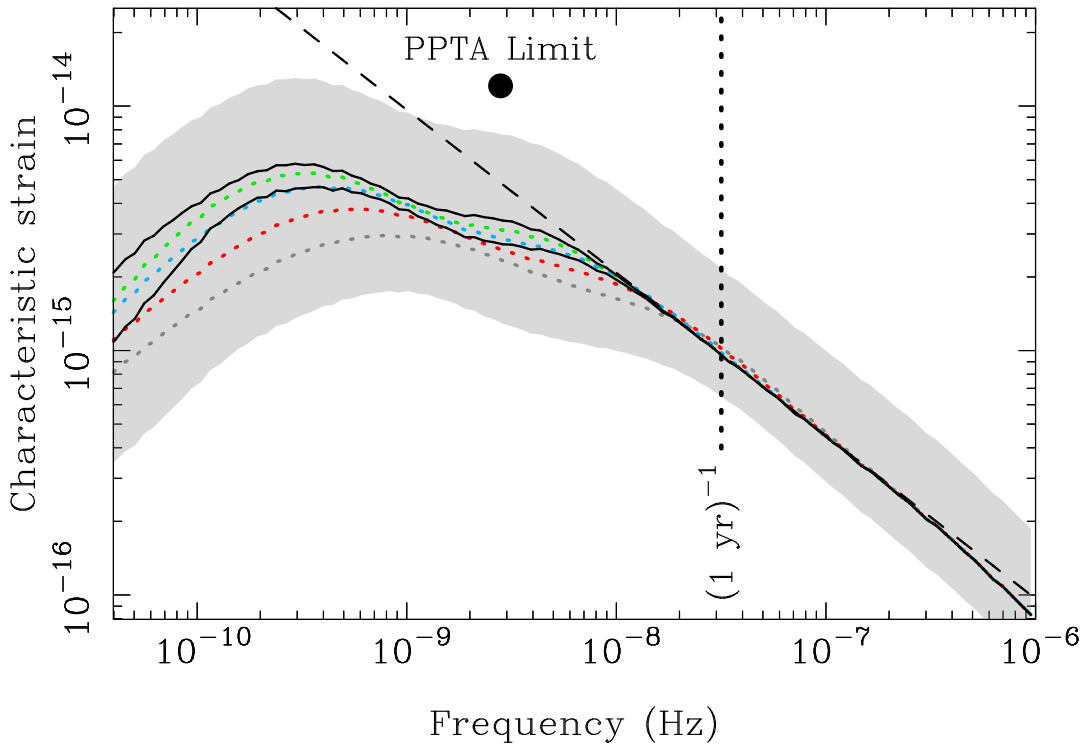

Figure 6.4: The four coloured, dotted curves are the characteristic strain spectra for the four w_0 cases I consider, also shown with the same colours in Figure 6.2. The upper solid black curve corresponds to a stellar density profile index of $\gamma = 1$, and the lower solid black curve corresponds to $\gamma = 2$; both are calculated assuming $w_0 = 0$, and so may be compared with the green (uppermost, $w_0 = 0$) dotted curve. The grey shaded area represents an approximate 68% confidence interval in the prediction of $h_c(f)$, given observational errors in the SMBH-bulge mass relation and the galaxy stellar mass function (a 0.4 dex range), allowing for the full range of w_0 values, and including a possible increase of 0.15 dex in the predictions if the binary SMBH population statistics were accurately specified. The characteristic strain spectrum calculated here in the circular, GW-driven case (and shown in Figures 6.2 and 6.3) is displayed as a black dashed line. The vertical dotted line indicates a frequency of $(1\,\mathrm{yr})^{-1}$, and the large black dot indicates the PPTA upper limit on the GWB at a frequency of 2.8×10^{-9} Hz.

I also include the assertion that the characteristic strain spectrum could be up to 0.15 dex larger than I calculate if the binary SMBH distribution were correctly specified.

It is clear that that there is relatively more uncertainty in the prediction at frequencies $f \lesssim 2\times10^{-8}$ Hz, where environmental interactions and binary eccentricities may affect the signal. I have also not included the uncertainty in the specific model for environment-driven binary SMBH evolution. As discussed above, the model I use may represent the maximum level of binary-environment coupling; other models may result in the characteristic strain spectrum being boosted at low frequencies relative to my prediction. For example, the model of Khan et al. (2012) suggests that the effects of environmental interactions may only be relevant for $f \lesssim 3.5 \times 10^{-9}$ Hz (also see Appendix A of this Chapter). I have also weighted each w_0-value equally, whereas it is possible that low-w_0 values are preferred over high-w_0 values.

In Figure 6.4, I also indicate the best upper bound on a stochastic, Gaussian GWB from binary SMBHs, published recently by the PPTA (Shannon et al. 2013). This upper bound corresponds to $\Omega_{\rm GW}(2.8\,{\rm nHz}) < 1.3 \times 10^{-9}$ with 95% confidence. While PTA bounds are traditionally shown as wedges (e.g., Sesana et al. 2008b, their Figure 13) on characteristic strain spectrum plots, Shannon et al. (2013) argued that their limit was applicable only at a single GW frequency. I hence display this limit as a single dot.

The prediction for the characteristic strain spectrum at a frequency of $f = (1\,{\rm yr})^{-1}$ of $6.5 \times 10^{-16} < h_c < 2.1 \times 10^{-15}$ (with approximately 68% confidence) is broadly consistent with previous results (e.g., Wyithe & Loeb 2003a; Sesana et al. 2008b) that considered the circular, GW-driven case. Indeed, for $f \gtrsim 2 \times 10^{-8}$ Hz, the predicted characteristic strain spectrum closely resembles the $h_c(f) \propto f^{-2/3}$ power law expected in the circular, GW-driven case, with the exception that for larger w_0 slightly more signal is present. The departure from the $f^{-2/3}$ power law at $f \sim 3\times10^{-7}$ Hz is caused by binary SMBHs radiating at these frequencies reaching their last stable orbits and not being included in the calculations (see, e.g., Wyithe & Loeb 2003a).

6.3.4 Summary of PTA implications

These results suggest a challenging future for attempts at detecting the GWB from binary SMBHs with PTAs. The frequency of optimal sensitivity for PTAs generally corresponds to the inverse of the characteristic observation time. Typical observation times of $5 - 30\,\mathrm{yr}$ (Manchester & IPTA 2013) imply that the properties of the GW signal at frequencies in the range $5 \times 10^{-10}\,\mathrm{Hz}$ to $10^{-8}\,\mathrm{Hz}$ are of primary importance for PTA work. The model I use in this paper implies that the GW characteristic strain spectrum may be reduced throughout this frequency range relative to the circular, GW-driven case (i.e., with respect to a $h_c(f) \propto f^{-2/3}$ power law). More generally, the gains in sensitivity to a GWB with observing time, estimated assuming $h_c(f) \propto f^{-2/3}$ (e.g., Siemens et al. 2013), are likely to be overestimated. Furthermore, it is possible that at low frequencies the increased contribution to the total GW signal of rare, massive binary SMBHs relative to the circular, GW case will cause the signal at these frequencies to be more non-Gaussian than suggested in Chapter 5.

However, this work requires significant refinement. It is clear from Figure 6.4 that the prediction for the characteristic strain spectrum at low frequencies is quite uncertain. Besides this uncertainty, the model I use in this paper for the coupling between binary SMBHs and stellar environments (Sesana et al. 2006; Sesana 2010) may in fact maximise the strength of this coupling. Also, it is possible that lower-eccentricity scenarios may be preferred over the higher-eccentricity scenarios. Both the above possibilities would result in the low-frequency parts of the characteristic strain spectrum being increased relative to my predictions. Further work on modelling the evolution of binary SMBH orbits in a variety of realistic galaxy merger scenarios is clearly necessary. This is of significant importance for predicting the strength of the GW signal from binary SMBHs in the PTA frequency band.

6.4 Predictions for GW bursts

6.4.1 The distribution of GW bursts

The prospect of detecting GW bursts with PTAs has been pursued recently by a number of authors (e.g., Finn & Lommen 2010; Pitkin 2012). GW burst detection algorithms generally contain few assumptions about the source properties, except

that they search for a strong signal confined to a short time-period. Here, I focus on the properties of GW bursts from eccentric binary SMBHs, and use the distribution of binary SMBHs, $D_\zeta[N(\zeta, z)]$, to predict the distribution of burst events.

It is necessary to form a definition of a GW burst from an eccentric binary SMBH in terms pulsar timing data products. Pulsar timing is the practice of measuring the ToAs of pulses from millisecond radio pulsars and fitting a physical model to these measurements. In Appendix B of this Chapter (§6.8), I describe how GWs from eccentric binaries affect pulsar timing measurements by inducing variations to ToAs. I present an expression for the rms deviation, $\sigma_R(t)$, of the ToA variations as a function of time caused by a given binary SMBH in Equation 6.19, averaged over all orientation parameters.

In Figure 6.5, I show the orbital phase, θ, the expected energy flux in GWs at the Earth and $\sigma_R(t)$ as functions of time for a binary SMBH with eccentricity 0.8 and orbital period at the Earth of 3.1 yr at the starting time, component masses $M_1 = 10^{10}\,M_\odot$ and $M_2 = 5 \times 10^9\,M_\odot$, and redshift 0.1. I also show the induced ToA variations corresponding to the GW metric perturbation at the Earth, $R(t)$, for arbitrary orientation parameters (binary inclination $i = 1\,\mathrm{rad}$, line of nodes orientation $\phi = 0.5\,\mathrm{rad}$). The orbital evolution of the binary was calculated using the work of Peters & Mathews (1963), and the energy flux at the Earth is averaged over binary inclination. The time-intervals considered to be GW bursts are highlighted in all panels of Figure 6.5. These 'bursts' correspond to the times of the largest change in the shortest amount of time in the ToA variations (see the top panel of Figure 6.5), and can be identified using $\sigma_R(t)$. I define the true burst amplitude, R_{burst}, to be twice the peak value of $\sigma_R(t)$, because that represents the expected peak-to-peak variation for a burst. The burst duration, T_{burst}, is the time-interval between peaks in $\sigma_R(t)$, represented by the widths of the shaded intervals in Figure 6.5. It is interesting that the GW bursts in the ToAs correspond to the motion of the binary through apastron, rather than periastron. This is because the GW-induced ToA variations are given by the time-integral of the GW amplitude as a function of time, as outlined in Appendix B.

The qualitative properties of $\sigma_R(t)$ in Figure 6.5 apply to binaries with any component masses, orbital period and eccentricity. That is, there are two peaks per rotation period, separated by less in orbital phase for more eccentric binaries, and separated by half an orbital phase for circular binaries. For a binary specified by ζ and

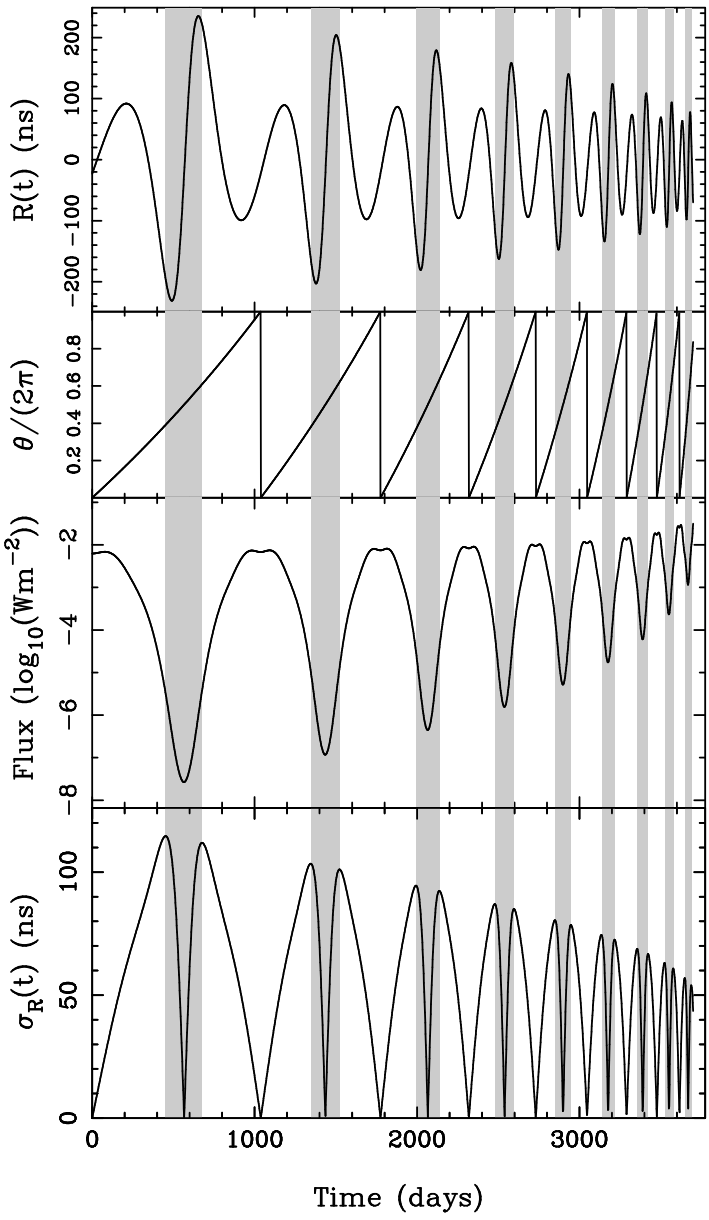

Figure 6.5: Properties of the GW signal and induced ToA variations for a binary SMBH with eccentricity 0.8, orbital period (at the Earth) of 3.1 yr at the starting time, component masses $M_1 = 10^{10}\,M_\odot$ and $M_2 = 5 \times 10^9\,M_\odot$, and redshift 0.1. The grey shadings in each panel represent the time-periods identified as GW bursts. The panels, from the top, are described in turn. *First panel:* the ToA variations corresponding to the metric perturbations at the Earth, for arbitrary orientation parameters, with the mean subtracted. *Second panel*: the orbital phase θ, measured from the common line of nodes and periapse. *Third panel*: the GW flux at the Earth. *Fourth panel*: the rms induced pulsar ToA variations, averaged over all orientation parameters. The time coordinate is measured at the Earth. Not all minima in the bottom curve are at a value of zero because of imperfect numerical sampling.

z, I integrate the equations for the evolution of the orbit (Peters & Mathews 1963) from zero orbital phase to numerically calculate $R_{\rm burst}$ as the mean of the first two peaks in $\sigma_R(t)$, and $T_{\rm burst}$ as the time-interval between peaks.

I use the distribution of binary SMBHs, $D_\zeta[N(\zeta, z)]$, to calculate the distribution of GW bursts. As described above, this distribution is specified as the number of binaries in discrete bins of width ΔM_1, ΔM_2, Δz, $\Delta f_{\rm orb}$ and $\Delta e_{\rm GW}$, where the eccentricity bin-widths depend on the other parameters. Scaling this distribution by the comoving volume shell between redshifts $z - \Delta z/2$ and $z + \Delta z/2$ specifies the number of observable binary SMBHs. For parameters at the midpoints of each bin, I calculate $R_{\rm burst}$ and $T_{\rm burst}$. I approximate the burst rate from binaries in a bin as the number of binaries divided by their period observed at the Earth, and record the expected number of bursts in a 10 yr time-span.

6.4.2 Results

Using the distributions of binary SMBHs for $w_0 = 0.1$ and $w_0 = 0.93$, I calculated the numbers of GW bursts for different values of the expected maximum level of ToA variations, $R_{\rm burst}$, and the duration, $T_{\rm burst}$. I depict the distributions of GW bursts in Figure 6.6 as the number of bursts per 10 yr observation time, N_{10}, per dex^2, in bins of 0.075 dex in $R_{\rm burst}$ and 0.05 dex in $T_{\rm burst}$. I only considered bursts with $R_{\rm burst} \geq 40$ ns and $0.1\,{\rm yr} \leq T_{\rm burst} \leq 10\,{\rm yr}$. An rms ToA variation of 40 ns corresponds to the best timing precisions currently achieved for millisecond radio pulsars (Osłowski et al. 2013; Hobbs 2013).

In total, I predict 0.06 bursts per 10 yr in the $w_0 = 0.93$ case with these strengths and durations, as compared with 0.12 bursts per 10 yr in the $w_0 = 0.1$ case. This difference in the total number of bursts is because of the smaller number of binary SMBHs that are expected to be observed if the population is generally more eccentric. However, I note that bursts from low-eccentricity binaries, which will dominate the burst population in the $w_0 = 0.1$ case, may be less detectable than bursts from high-eccentricity binaries. There are proportionally more short-duration bursts in the $w_0 = 0.93$ case than in the $w_0 = 0.1$ case, because larger binary eccentricities result in shorter bursts.

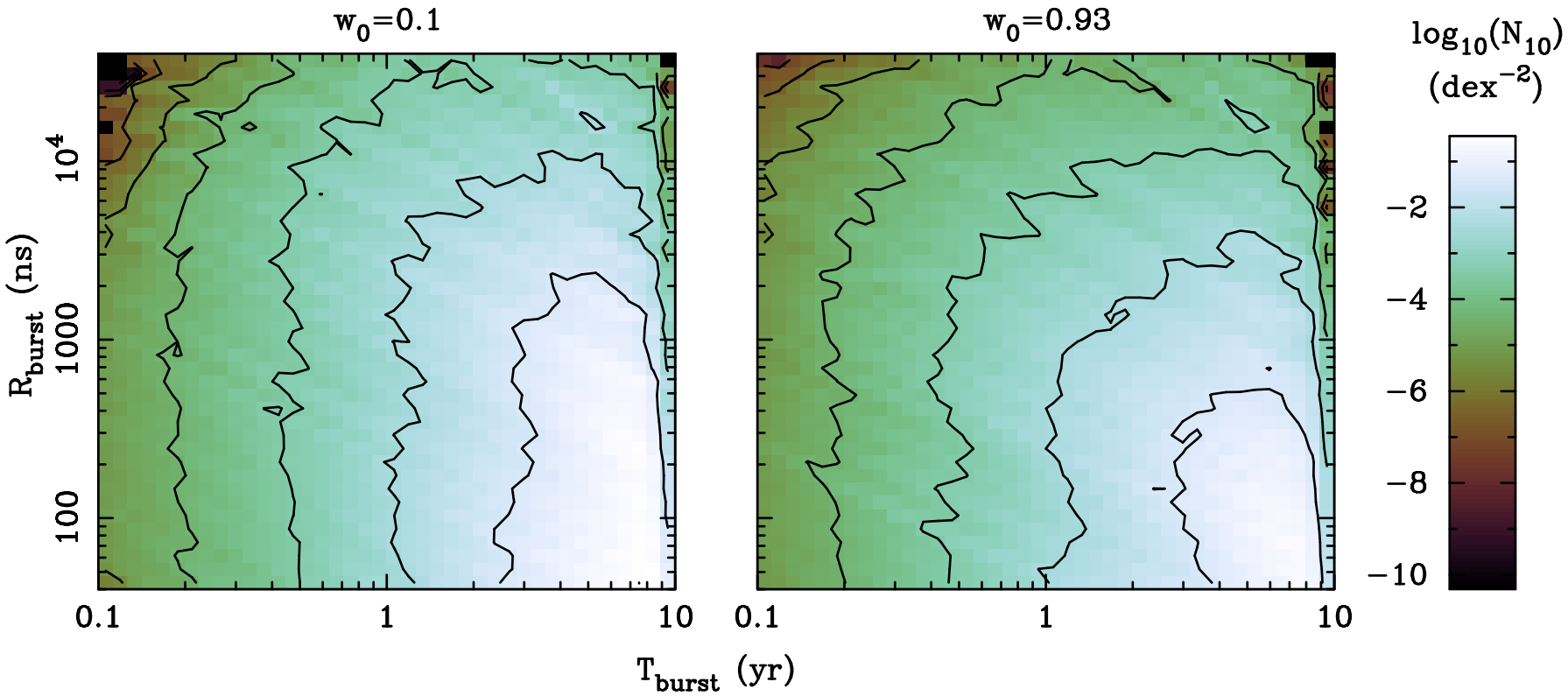

Figure 6.6: Illustrations of the distributions of GW bursts in the $w_0 = 0.1$ (left) and $w_0 = 0.93$ (right) cases. The shading represents the expected number of bursts in a 10 yr time-span, N_{10}, per dex^2. The distributions are binned over 0.05 dex in duration and 0.075 dex in amplitude. The contours connect regions at intervals of factors of 10 below the peak.

In both cases, the burst distribution is quite heavily skewed towards long bursts, with approximately a factor of 100 more bursts expected with $\sim 8\,\mathrm{yr}$ durations than with $\sim 1\,\mathrm{yr}$ durations. There are also fewer bursts with durations longer than $\sim 8\,\mathrm{yr}$ in both cases. This typical burst duration corresponds to binaries with separations where GW-driven evolution is equivalent to evolution driven by stellar environments.

The typical combined masses of the binary SMBHs that produce GW bursts are $\sim 10^{10}\,M_{\odot}$. In Figure 6.7, I show the distributions of the combined masses of all binary SMBHs producing the bursts in the distributions in Figure 6.6. The distributions are similar in shape, although the distribution for $w_0 = 0.93$ includes relatively more high-mass binaries than the distribution for $w_0 = 0.1$. This is because, in the $w_0 = 0.93$ case, lower-mass binaries are less likely to be able to produce strong GW bursts because they are likely to be more eccentric. More eccentric binaries of a given mass and orbital period produce typically weaker bursts (see Equation 6.19).

In summary, I find:

1. For bursts with durations between 0.1 yr and 10 yr, and with expected maximum ToA variations of $> 40\,\mathrm{ns}$, I predict between 0.06 and 0.12 bursts per 10 yr observation, with lower burst rates corresponding to higher-eccentricity binary SMBH populations.

2. Higher-eccentricity binary populations result in relatively more shorter duration bursts than lower-eccentricity populations.

3. However, the burst rate decreases by a factor of 10 per $\sim 0.4\,\mathrm{dex}$ below a duration of $\sim 8\,\mathrm{yr}$. This also appears to be the most likely duration, with few bursts longer than 8 yr expected.

4. The burst rate decreases by a factor of 10 per $\sim 0.8\,\mathrm{dex}$ increase in amplitude.

Various uncertainties discussed in §6.3.3 also apply to these calculations. The uncertainty in the galaxy merger rate will also directly correspond to the uncertainty in the GW burst rate (i.e., 0.3 dex). Given that the high-end power-law logarithmic slope of the SMBH mass function in the G11 model is ~ -2, the 0.2 dex uncertainty in the SMBH masses will, to first order, correspond to an uncertainty of 0.4 dex in the merger rate. Therefore, the 1σ uncertainty in the burst rate from the model is approximately 0.5 dex.

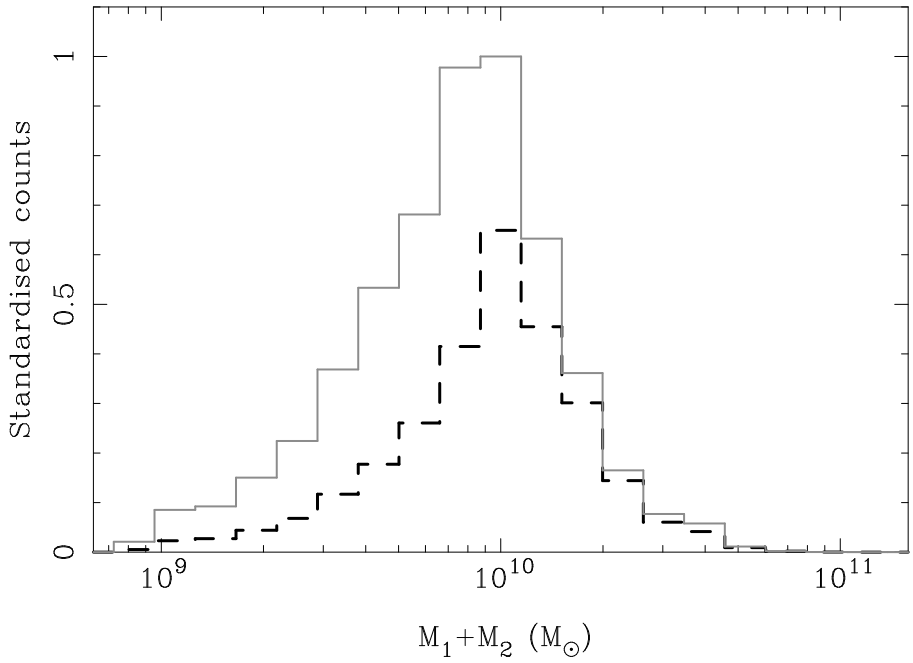

Figure 6.7: The distributions of combined masses (M_1+M_2) of binary SMBHs contributing to the GW burst distributions presented in Figure 6.6. The solid grey histogram corresponds to the $w_0 = 0.1$ case, and the dashed black histogram corresponds to the $w_0 = 0.93$ case. Both histograms are normalised to the peak of the $w_0 = 0.1$ histogram.

To my knowledge, only one study has attempted to predict the properties of GW bursts from a population of eccentric binary SMBHs (Liu et al. 2012). While their modelling methods and definition of GW bursts differ substantially from the present work, I agree with these authors that it is unlikely that current PTAs will be able to detect GW bursts from binary SMBHs. The rarity of short GW bursts (lasting around 1 yr) from binary SMBHs suggests that high-cadence PTA observations targeting such bursts are not well motivated.

6.5 Summary of results

I have used a semi-analytic model for galaxy and SMBH formation and evolution (G11) implemented in the Millennium simulation (Springel et al. 2005), augmented with a model for the evolution of binary SMBHs within fixed stellar backgrounds (Sesana et al. 2006; Sesana 2010), to predict the properties of low-frequency GWs from binary SMBHs. I specify the form of a phenomenological distribution of initial binary eccentricities, and consider a selection of cases with differing levels of typical binary eccentricity.

My quantitative results are uncertain due to a variety of factors. The range of initial binary eccentricity distributions that I consider corresponds to a 0.4 dex variation in the characteristic strain spectrum at low frequencies. Moreover, uncertainties in the tuning of the G11 model provide another 0.2 dex of uncertainty in the spectrum at all frequencies. There is also uncertainty in the estimate of the binary SMBH distribution predicted by the G11 model, in particular for the most massive binaries. Finally, while the G11 model is likely to provide a satisfactory representation of the merger rate of massive, low-redshift galaxies, the binary evolution model that I use may overestimate binary hardening caused by stellar interactions.

My specific findings are as follows:

1. The expected characteristic strain spectrum of the GWB from binary SMBHs will turn over from the standard $h_c(f) \propto f^{-2/3}$ power law at a frequency up to $\sim 10^{-8}$ Hz. The turn-over frequency depends on the efficiency of stellar interactions in extracting energy and angular momentum from binary SMBHs, as well as the typical binary eccentricities at formation.

2. The nature of the spectrum at frequencies below the turn-over frequency is extremely uncertain, and depends on the numbers of massive ($M_1 + M_2 \geq 10^{10} M_\odot$) binaries and on binary eccentricities. The most massive binary SMBHs predominantly produce the lowest-frequency parts of the spectrum, and their numbers depend strongly on the efficiency of the energy and angular momentum transfer to their environments. The spectrum will be attenuated if binaries with typically larger eccentricities are present.

3. The most massive eccentric binaries will very rarely produce GW bursts detectable in pulsar timing data. A larger-eccentricity binary population will produce fewer bursts that are typically shorter and weaker.

I emphasise a set of key implications of this work for PTAs:

1. Given the expected low-frequency turn-over in the GW characteristic strain spectrum, along with the large uncertainty in the signal at these frequencies, the increase with time of PTA sensitivities to a GWB from binary SMBHs will not be as strong as currently thought.

2. Short-duration, strong GW bursts from eccentric binary SMBHs are unlikely to occur during typical PTA dataset lengths.

3. PTA data analysts cannot assume $h_c(f) \propto f^{-2/3}$ when searching for a GWB from binary SMBHs. Indeed, *model-independent* searches cannot assume any particular spectral shape.

4. *Model-dependent* searches and constraints need to carefully account for the uncertainty in model predictions.

6.6 Conclusions

In this Chapter, I predicted both the GWB characteristic strain spectrum and the distribution of strong GW bursts from eccentric binaries. At a GW frequency of $(1\,\mathrm{yr})^{-1}$, I found a characteristic strain of $6.5 \times 10^{-16} < h_c < 2.1 \times 10^{-15}$ with approximately 68% confidence.

Accelerated binary evolution driven by three-body stellar interactions causes the characteristic strain spectrum to be diminished with respect to a $h_c(f) \propto f^{-2/3}$ power-law at $f \lesssim 2 \times 10^{-8}\,\mathrm{Hz}$. At these low frequencies, the signal is further attenuated if binary SMBHs are typically more eccentric at formation. The low-frequency signal may be dominated by a few binaries with combined masses $(M_1 + M_2)$ greater than $10^{10}\,M_\odot$, to a larger extent than predicted in the circular, GW-driven case (Ravi et al. 2012). Numerous uncertainties, however, affect my results. These include observational uncertainties in parameters of the model, and theoretical uncertainties in the efficiency of coupling between binary SMBHs and their environments.

I also expect between 0.06 and 0.12 GW bursts that produce >40 ns amplitude ToA variations over a 10 yr observation time. Larger typical binary eccentricities at formation will result in fewer events than if binaries are less eccentric at formation. These bursts are caused by binary SMBHs with combined masses of $\sim 10^{10}\,M_\odot$, and typically last $\sim 8\,\mathrm{yr}$. Shorter, stronger bursts are significantly less likely, as are longer bursts.

Upcoming radio telescopes with extremely large collecting areas, such as the Five hundred metre Aperture Spherical Telescope (FAST; Li et al. 2013; Hobbs et al. 2014) and the Square Kilometre Array (SKA; Cordes et al. 2004; Lazio 2013) are likely to significantly expand the sample of pulsars with sufficient timing precision for GW detection as compared to current instruments. PTAs formed with FAST and the SKA will hence be sensitive to a stochastic GW sig-

nal at much higher frequencies than current PTAs, which is desirable given the results I present.

The mechanism by which binary SMBHs are driven to the GW-dominated regime must involve some form of binary-environment coupling. Hence, independent of the exact model, *there will always be some low-frequency attenuation of the GW signal relative to the circular, GW-driven binary case.* My results indicate that this attenuation occurs within the PTA frequency band. However, the strength of the binary-environment coupling is quite uncertain, and I urge future work on this topic. Finally, constraining or measuring the spectrum of the GWB at a number of frequencies would provide an excellent test of models for the binary SMBH population of the Universe.

6.7 Appendix A: Testing the binary SMBH evolution model

In the main text, I use the results of Sesana et al. (2006) and Sesana (2010) (hereafter collectively referred to as S06 in this section) to model the evolution of binary SMBHs in fixed stellar backgrounds for separations less than a_h (see Equation 3.6). S06 numerically solved three-body scattering problems for binary SMBHs interacting with stars on radial, intersecting orbits drawn from a spherically-symmetric, fixed distribution, and provided fitting formulae for the binary hardening and eccentricity growth rates as functions of binary properties. I use these fitting formulae to evolve binary SMBH orbits as described above.

Here, I compare this method of evolving binary SMBHs with recent N-body simulations of binary SMBH evolution in merging galaxies of various mass ratios and stellar density distributions (Khan et al. 2012, hereafter K12 in this section). K12 simulated the mergers of spherical galaxies with various mass ratios, power-law stellar cusp density profiles with various indices, with typical approach trajectories from cosmological simulations. The SMBHs were traced until separations close to, and in some cases beyond, where the GW-driven orbital decay dominated the orbital decay caused by three-body stellar interactions. By extrapolating the binary orbits assuming constant stellar-driven hardening rates and eccentricities, K12 estimated binary semi-major axes, $a_{\rm K12}$, below which GW-driven evolution dominated.

Table 6.1: Comparison between decoupling times for S06 and K12 models.

Model	γ	q	$\frac{a_{\rm S06}}{a_{\rm K12}}$	$\frac{a_{\rm S06}}{a_{\rm K12}}$ (fixed e)
A1	0.5	0.1	0.51	0.47
A2	0.5	0.25	0.72	0.53
A3	0.5	0.5	0.62	0.57
A4	0.5	1.0	0.68	0.65
B1	1.0	0.1	0.58	0.36
B2	1.0	0.1	0.74	0.38
B3	1.0	0.1	0.87	0.41
B4	1.0	0.1	0.93	0.47
D1	1.75	0.1	0.72	0.41
D2	1.75	0.1	0.68	0.44
D3	1.75	0.1	0.62	0.48
D4	1.75	0.1	0.63	0.51

The accuracy of these extrapolations was confirmed using a selection of simulations including post-Newtonian corrections to the binary SMBH orbits.

Here, I take the final eccentricities and semi-major axes of the binaries in each of the scenarios considered by K12 for which $a_{\rm K12}$ was estimated, and evolve the binaries using the binary evolution model of S06 to estimate an equivalent quantity to $a_{\rm K12}$, $a_{\rm S06}$. I list the ratios $a_{\rm S06}/a_{\rm K12}$ in Table 6.1 both without (column 4) and with (column 5) the binary eccentricity held fixed for each relevant scenario of K12. The cusp density profile indices, γ, and the galaxy and SMBH mass ratios, q, are given in columns 2 and 3 respectively.

I find $0.1 < \frac{a_{\rm S06}}{a_{\rm K12}} < 1$ in all cases. This implies that the S06 model that I use in this work has stronger stellar-driven binary evolution than the K12 model. I also find smaller ratios $\frac{a_{\rm S06}}{a_{\rm K12}}$ when I hold the binary eccentricities fixed. This is because the binary eccentricities invariably grow when allowed to evolve, and lower eccentricities imply smaller GW-driven hardening rates. While an intuitive explanation of the difference between the S06 and K12 models is difficult to attain, the K12 work involves a more sophisticated, and possibly more realistic, treatment of the distribution of stellar orbits in the cores of merged galaxies than the S06 work. Given $f_{\rm orb} \propto a^{-3/2}$, a difference of a factor of two in the semi-major axes at which binary SMBH evolution begins to be GW-dominated corresponds to a difference of a factor of $2^{3/2} (\sim 0.45\,{\rm dex})$ in orbital frequency.

6.8 Appendix B: GW bursts and pulsar ToAs

Here, I provide a mathematical description of GW bursts from eccentric binary SMBHs. The spatial metric perturbation tensor, or GW strain, corresponding to a binary SMBH was defined by Wahlquist (1987) to lowest order in the slow-motion, far-field regime using the quadrupole formula. This tensor, h_{ij}, can be written as (Wahlquist 1987)

$$h_{ij} = \sum_{S=+,\times} h^S e^S_{ij}, \tag{6.9}$$

where e^+_{ij} and $e^\times_{ij}$ are the 'plus' and 'cross' polarisation tensors respectively, and h^+ and $h^\times$ are the time-varying polarisation amplitudes, which depend on the orbital phase θ (which is a function of time), the value of θ at the line of nodes (θ_n), the value of θ at periastron (θ_p), the orientation of the line of nodes (ϕ), the binary inclination (i), and a factor

$$A_g = \frac{4(GM_C)^{5/3}}{c^4 D(z)(1-e^2_{\rm GW})}(2\pi f_{\rm orb})^{2/3}, \tag{6.10}$$

where $D(z)$ is the comoving coordinate distance to redshift z and $M_C = (M_1 M_2)^{3/5}(M_1+M_2)^{-1/5}$ is the binary chirp mass.

GWs at the Earth and at a pulsar cause a fractional shift in the observed pulsar rotation frequency. Here, I neglect the effects of GWs at the pulsar, because, as outlined in, e.g., Finn & Lommen (2010), GW bursts will generally affect pulsar timing data at vastly different times for different pulsars. This means that an approach that seeks to detect GW bursts by observing correlated effects in multiple pulsar datasets will only need to consider the effects of GWs at the Earth. For a pulsar with location defined by the unit direction tensor p^i, the observed fractional pulsar rotation frequency shift is given by (Wahlquist 1987; Hobbs et al. 2009)

$$\frac{\delta\nu_p}{\nu_p} = \frac{-p^i p^j h_{ij}}{2(1+\mu)}, \tag{6.11}$$

where μ is the cosine of the angle between the pulsar and GW source directions, and I follow the Einstein summation convention over the tensor indices.

Fractional shifts in ν_p will cause cumulative variations in ToAs. That is,

$$R(t) = \int_0^t \frac{\delta\nu_p}{\nu_p} dt, \tag{6.12}$$

where $R(t)$ is the ToA variation at a time t. In order to calculate the average duration and strength of a GW burst from a binary SMBH as manifested in $R(t)$, I need to calculate the variance

$$\sigma_R^2(t) = \langle R^2(t) \rangle_{\theta_n, \theta_p, \phi, i, \alpha, \delta} \tag{6.13}$$

where the angle brackets signify averaging over the subscripted quantities, and α and δ are the right ascension and declination of the pulsar assuming that the GW source is located along the z-axis. To simplify this, I set $\theta_n = \theta_p = 0$, and express $R(t)$ as

$$R(t) = \sum_{S=+,\times} R^S(t) G^S, \tag{6.14}$$

where

$$R^{+,\times}(t) = \int_0^t h^{+,\times} dt \tag{6.15}$$

and

$$G^{+,\times} = -\frac{p^i p^j e_{ij}^{+,\times}}{2(1+\mu)}. \tag{6.16}$$

The linear independence of the polarisation tensors implies that

$$\sigma_R^2(t) = \sum_{S=+,\times} \langle (R^S(t))^2 \rangle_{\phi, i} \langle (G^S)^2 \rangle_{\alpha, \delta}. \tag{6.17}$$

I find that

$$\langle (G^{+,\times})^2 \rangle_{\alpha, \delta} = \frac{1}{6}. \tag{6.18}$$

Then,

$$\sigma_R^2(t) = \frac{A_g^2 (1 - e_{\rm GW}^2)^3 \sin^2\theta}{720 f_{\rm orb}^2 \pi^3 (1 + e_{\rm GW} \cos\theta)^2} (3(8 + e_{\rm GW}^2) + 16 e_{\rm GW} \cos\theta + 4\cos(2\theta)) \tag{6.19}$$

To summarise, Equation 6.19 gives the variance of the ToA variations caused by GWs from an individual eccentric binary SMBH at the Earth, averaged over all orientation parameters.

Chapter 7

Prospects for GW detection and SMBH astrophysics with PTAs

Large-area sky surveys show that massive galaxies undergo at least one major merger in a Hubble time. Ongoing pulsar timing array (PTA) experiments are aimed at measuring the gravitational wave (GW) emission from binary supermassive black holes (SMBHs) at the centres of galaxy merger remnants. In this Chapter, using the latest observational estimates for a range of galaxy properties and scaling relations, I predict the amplitude of the GW background generated by the binary SMBH population. I also predict the numbers of individual binary SMBH GW sources. I find the characteristic strain amplitude of the GW background to lie in the range $5.1 \times 10^{-16} < \mathrm{A_{yr}} < 2.4 \times 10^{-15}$ at a frequency of $(1\,\mathrm{yr})^{-1}$, with 95% confidence. Higher values within this range, which correspond to the more commonly preferred choice of galaxy merger timescale, will fall within the expected sensitivity ranges of existing PTA projects in the next few years. In contrast, I find that a PTA consisting of at least 100 pulsars observed with next-generation radio telescopes will be required to detect continuous-wave GWs from binary SMBHs. I further suggest that GW memory bursts from coalescing SMBH pairs are not viable sources for PTAs. Both the GW background and individual GW source counts are dominated by binaries formed in mergers between early-type galaxies of masses $\gtrsim 5 \times 10^{10}\mathrm{M}_{\odot}$ at redshifts $\lesssim 1.5$. Uncertainties in the galaxy merger timescale and the SMBH mass – galaxy bulge mass rela-

tion dominate the uncertainty in the predictions. This Chapter describes my original work, which has largely been published (Ravi et al. 2015).

7.1 Introduction

I have so far quantified the summed GW signal (the GWB) from binary SMBHs throughout the Universe using a physically-motivated model for the SMBH-SMBH coalescence rate. This model was based on analytic prescriptions for the evolution of the baryon content of the Universe (Guo et al. 2011) implemented within dark matter halo merger trees from the Millennium simulation (Springel et al. 2005). In my analysis of the statistics of GW-induced variations to pulsar timing data in Chapter 5, I assumed that all binary SMBHs emitting GWs in the PTA frequency band ($10^{-9} - 10^{-7}$ Hz) are in circular orbits, losing energy and angular momentum only to gravitational radiation. Upon relaxing this assumption in Chapter 6, I found that the expected GWB characteristic strain spectrum is likely attenuated at low frequencies, potentially up to 10^{-8} Hz, from the power law spectral form in Equation 5.1 ($h_c(f) = A_{\rm yr}(f/f_{\rm yr})^{-2/3}$). Here, I approach the problem of characterising the SMBH-SMBH coalescence rate from an empirical perspective, in order to determine the possible range of GW signals in the PTA frequency band.

Stellar- or gas-dynamical evidence exists for SMBHs at the centres of 87 nearby galaxies at the time of writing (Kormendy & Ho 2013). Phenomenological models of the buildup of the cosmological mass density in SMBHs during luminous quasar phases for redshifts $z < 5$ (e.g., Yu & Tremaine 2002; Shankar et al. 2013) indicate short quasar lifetimes. This, considered together with local correlations between SMBH masses, $M_\bullet$, and, for example, galaxy bulge masses $M_{\rm bul}$ (e.g., Kormendy & Ho 2013), suggests that all massive galaxies ($M_* > 10^{10} M_\odot$) which formed since the $z \sim 2$ peak of quasar activity host SMBHs (see also Miller et al. 2014).

In the context of hierarchical structure formation, mergers are integral to the formation histories of massive galaxies, and evidence for interacting galaxies is seen across most of cosmic time (Barnes & Hernquist 1992). Multiple SMBHs are expected to be found in galaxy merger products. Indeed, pairs of active galactic nuclei (AGN) are observed in galaxies that are in the late stages of mergers (Merritt & Milosavljević 2005), with projected separations as small as 7 pc

(Rodriguez et al. 2006). As described in §3.1, the central SMBHs in a pair of merging galaxies are likely to sink in the merger remnant potential well through dynamical friction and form a gravitationally-bound binary, which coalesces on a timescale much less than the galaxy merger timescale. A few candidate binary SMBHs have been identified (e.g., Valtonen et al. 2008; Boroson & Lauer 2009; Eracleous et al. 2012).

The existence of a large cosmological population of binary SMBHs, some of which emit GWs in the PTA frequency band, is thus inferred. The summed signals from the population of binary SMBHs may be collectively modelled as a GWB. GWs from individual binaries are potentially detectable as continuous-wave (CW) sources (see Chapter 5, and also Sesana et al. 2009). Coalescing pairs of SMBHs also emit GW 'memory' bursts, which are abrupt, propagating metric changes that also affect pulsar timing measurements (e.g., Madison et al. 2014). Results from the most sensitive searches for and constraints on these three types of GW signal were summarised in Chapter 4.

Predictions for GW signals from binary SMBHs are based either on physical models for galaxy formation and evolution which predict the binary SMBH population, or on directly-observed quantities such as the galaxy merger rate and SMBH-galaxy scaling relations. The former approach, as demonstrated in Chapters 5 and 6, typically combines dark matter halo merger rates in the cold dark matter paradigm, analytic or numerical estimates of galaxy merger and binary SMBH formation timescales, prescriptions for the cosmic evolution of the galaxy and SMBH population and the assumption of GW-driven binary orbital evolution (see also Wyithe & Loeb 2003a; Enoki et al. 2004; Sesana et al. 2008b; Kulier et al. 2013). Once tuned to reproduce observables such as local SMBH-galaxy relations, the galaxy stellar mass function (GSMF) and colour distribution, and the quasar luminosity function, these models result in estimates of $A_{\rm yr}$ in the range $10^{-16} - 2 \times 10^{-15}$. The exact value depends on, for example, the assumed cosmological parameters, galaxy merger timescales and the specific models for SMBH formation and growth. Models for the binary SMBH population which more directly incorporate observational information (Jaffe & Backer 2003; Sesana 2013b; McWilliams et al. 2014) result in similar values of $A_{\rm yr}$. While individual binary SMBHs may be viable CW sources of GWs for PTAs, as well as viable sources of memory bursts (van Haasteren & Levin 2010; Cordes & Jenet 2012;

Madison et al. 2014), quantitative predictions of source counts have only been calculated for CW sources using theoretical galaxy formation models (Sesana et al. 2009).

A recent study by Sesana (2013b), hereafter in this Chapter S13, combined a plethora of observational estimates of the merger rate of massive galaxies, the GSMF and local SMBH-galaxy relations to derive a range of possible GWB amplitudes. However, some of the observational quantities included in the S13 study do not represent the best current knowledge. Furthermore, some relevant uncertainties, such as in the possible redshift-evolution of the SMBH-galaxy relations, were not accounted for by S13, and no predictions for individual GW sources were made. Besides addressing these issues, this Chapter builds on previous studies in the following ways:

- I quantify the impacts of different observational uncertainties on the amplitude of the GWB generated by binary SMBHs. I focus in particular on aspects of a new model for the binary SMBH population for which little observational information currently exists.

- I highlight the redshifts, masses and types of merging galaxies which result in binary SMBHs that dominate the GWB amplitude.

- I provide new, observations-based predictions for the counts of individual GW sources. I further present the first estimates for the expected numbers of detectable individual GW sources, given different PTA configurations, that are robust with respect to pulsar parameter fitting.

In §7.2, I outline the model, and present my results in §7.3. I state the key implications of this work for PTAs in §7.4. Finally, I discuss these results in §7.5, and summarise my conclusions in §7.6. Throughout this Chapter, I adopt a concordance cosmology based on results from the *Planck* satellite (Planck Collaboration et al. 2013), including $H_0 = 67.8\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $\Omega_\Lambda = 0.692$ and $\Omega_M = 0.308$.

7.2 An empirical model for GWs from binary SMBHs

7.2.1 The SMBH-SMBH coalescence rate

As demonstrated in the previous Chapters, the cosmological population of binary SMBHs emitting GWs in the PTA frequency band can be characterised using the SMBH-SMBH coalescence rate. The numbers of binary SMBHs in different orbits are related to the coalescence rate through a continuity equation (e.g., Equations 3.52 and 6.4) that includes assumptions about the rate of binary SMBH orbital evolution.

I assume that the SMBH-SMBH coalescence rate is equivalent to the galaxy merger rate. This is justified because, as outlined in §3.1.3, massive galaxy mergers are typically completed within a few galaxy dynamical times, whereas the timescales for two SMBHs to form gravitationally-bound binaries and then coalesce through losses of energy and angular momentum to their environments and GWs are much shorter. I neglect systems of more than two gravitationally-interacting SMBHs resulting from multiple galaxy mergers, because these are expected to be rare for the high mass ratio ($\mu_* > 1/3$) mergers between massive ($M_* > 10^{10} M_\odot$) galaxies that I consider in this Chapter. I further assume that each galaxy contains a central SMBH with a mass related to the galaxy bulge mass. I use measured quantities to determine the all-sky coalescence rate of pairs of SMBHs. For each quantity, I define a fiducial prescription, and also describe the possible ranges over which the prescription can vary.

Similarly to S13, I express the galaxy merger rate as

$$\Phi_{\rm mrg}(M_*, \mu_*, z) = \frac{d^4 N_{\rm mrg}}{d\log(M_*) d\log(\mu_*) dz dt} \tag{7.1}$$

$$= \frac{1}{\Gamma} \frac{dt_p}{dt} \frac{d^2 N_{\rm gal}}{d\log(M_*) dz} \frac{dP}{d\log(\mu_*)}\Big|_{M_*} \tag{7.2}$$

where $N_{\rm mrg}$ is the number of mergers between two galaxies of combined stellar mass $M_*(1+\mu_*)$, μ_* is the ratio between the smaller and larger galaxy stellar masses and z is the cosmological redshift. The merger rate, $\Phi_{\rm mrg}$, is defined as the number of mergers per units M_*, μ_*, z and observer time t. In Equation 7.2, $N_{\rm gal}$ is the number

of galaxies across the entire sky with a given M_* at a given z. This is related to the standard GSMF, $\Phi_*(M_*,\, z)$, as

$$\frac{d^2 N_{\rm gal}}{d\log(M_*) dz} = \Phi_* \frac{4\pi d^2 V_c}{d\Omega dz}, \tag{7.3}$$

where $\frac{4\pi d^2 V_c}{d\Omega dz}$ is the sky-integrated comoving volume shell between redshifts z and $z + dz$. In Equation 7.3, $\frac{dP}{d\log(\mu_*)}\big|_{M_*}$ is the probability density function for a galaxy merger event with mass M_* at redshift z having a mass ratio μ_*, $\Gamma(M_*,\, z) = \left(\frac{dn_{\rm mrg}}{dt_p}\right)^{-1}$ is the average proper time between major mergers for a galaxy with a mass M_* at redshift z and $\frac{dt_p}{dt} = (1+z)^{-1}$. Also, $\frac{dn_{\rm mrg}}{dt_p}\big|_{M_*,\, z}$ is the number of mergers, $n_{\rm mrg}$, per unit proper time, t_p, for a single galaxy with a mass M_* at redshift z.

In order to convert galaxy stellar masses to bulge masses ($M_{\rm bul}$) I distinguish between quiescent, red-sequence early-type galaxies and star-forming, blue-cloud late-type galaxies. I write the total GSMF as a sum of the GSMFs of early- ($\Phi_{*,\,\rm early}$) and late-type ($\Phi_{*,\,\rm late}$) galaxies:

$$\Phi_*(M_*,\, z) = \Phi_{*,\,\rm early} + \Phi_{*,\,\rm late}. \tag{7.4}$$

I relate M_* to $M_{\rm bul}$ for early- and late-type galaxies using a scheme described below.

To convert between $M_{\rm bul}$ and the central SMBH masses ($M_\bullet$) I use the widely-known $M_\bullet - M_{\rm bul}$ relation (Kormendy & Ho 2013; Scott et al. 2013). In contrast to S13 and Chapter 2, I express this relation as

$$\frac{dP}{d\log M_\bullet} = \mathcal{N}(\alpha + \beta \log M_{\rm bul}, \epsilon^2) \tag{7.5}$$

where $\mathcal{N}(\mu,\, \sigma^2)$ denotes a normal probability density function with centre μ and variance σ^2, and α, β and the intrinsic scatter, ϵ, are observationally-determined constants. It is important to account for intrinsic scatter in the $M_\bullet - M_{\rm bul}$ relation when inferring the SMBH mass function from the bulge mass function (e.g., Aller & Richstone 2002), because to not do so would lead to the SMBH mass function being underestimated.

7.2.1.1 The times between galaxy mergers

Observational estimates of $\Gamma(M_*,\, z)$ require knowledge of the fraction of galaxies, $f_{\rm gm}$, within a mass-complete sample at a given redshift that are undergoing mergers, and the proper time τ_m during which merger events can be observationally identified (for a review, see Conselice 2014). Then, $\Gamma(M_*,\, z) = \tau_m / f_{\rm gm}$. In this work, I focus on major mergers with stellar mass ratios $\mu_* \geq 1/3$, because these systems are likely to dominate the GW signal (e.g., Sesana et al. 2004, S13).

I consider three recent measurements of $f_{\rm gm}$ for major mergers at different redshifts in wide-area galaxy surveys, which are largely complete for galaxy stellar masses $M_* > 10^{10} M_\odot$. These three studies fit their data to the function $f_{\rm gm} = a_{\rm gm}(1+z)^{b_{\rm gm}}$, where $a_{\rm gm}$ and $b_{\rm gm}$ are free parameters.

- Conselice et al. (2009) used structural analyses of concentration, asymmetry and clumpiness (the 'CAS' parameters) to identify merging systems among $\sim$ 22000 galaxies in the COSMOS and Extended Groth Strip surveys with $M_* > 10^{10} M_\odot$ at $z < 1.2$. This technique is sensitive to major mergers in particular, with mass ratios $\mu_* \gtrsim 1/3$ (Conselice 2003). Conselice et al. (2009) found $f_{\rm gm} = (0.022 \pm 0.006)(1+z)^{1.6 \pm 0.6}$.

- Xu et al. (2012) counted galaxy pairs with projected separations between $5\,h^{-1}\,$kpc and $20\,h^{-1}\,$kpc from the COSMOS survey to estimate $f_{\rm gm}$ for $z < 1$ and $\mu_* > 0.4$. They scaled their results to include galaxy pairs for all $\mu_* \geq 1/3$ using the argument that $f_{\rm gm}$ is inversely proportional to the logarithm of minimum mass ratio of the observed galaxy pair sample. Xu et al. (2012) found $f_{\rm gm} = (0.013 \pm 0.001)(1+z)^{2.2 \pm 0.2}$.

- Both the above works may suffer from incorporating small galaxy samples at low redshifts. This issue was addressed by Robotham et al. (2014) using a large sample of galaxy pairs from the Galaxy and Mass Assembly (GAMA) survey in the redshift interval $0.05 < z < 0.2$. When standardised to the same projected separation, galaxy mass and mass ratio windows as Xu et al. (2012), they found a substantially higher value of $f_{\rm gm}$ at these redshifts. By combining their results with all recent measurements of $f_{\rm gm}$ at redshifts up to 1.2, and normalising to the same projected pair separations of Xu et al. (2012), Robotham et al. (2014) found $f_{\rm gm} = (0.021 \pm 0.001)(1+z)^{1.53 \pm 0.08}$.

Other works have estimated $f_{\rm gm}$ with varying levels of accuracy. S13 included results from the galaxy pair studies of Bundy et al. (2009), de Ravel et al. (2009) and López-Sanjuan et al. (2012). However, Bundy et al. (2009) and de Ravel et al. (2009) had significantly smaller pair samples than were utilised by either Xu et al. (2012) or Robotham et al. (2014), and López-Sanjuan et al. (2012) only considered major mergers of galaxies with $M_* > 10^{11} M_\odot$.

In the absence of observational estimates of the galaxy merger timescale (τ_m) used in calculating $\Gamma(M_*,\, z)$ from different measurements of $f_{\rm gm}$, I make use of theoretical predictions. However, the range of possible predictions spans a factor of three. Kitzbichler & White (2008) used a mock galaxy catalogue from a semi-analytic model implemented within the Millennium simulation (Springel et al. 2005) to estimate τ_m for galaxies with different masses at different stages of merging, assuming circular galaxy orbits and angular momentum loss through dynamical friction. However, a suite of hydrodynamic simulations of galaxy mergers conducted by Lotz et al. (2008) and Lotz et al. (2010), hereafter collectively L08 in this Chapter, resulted in significantly shorter merger timescales. While some authors (e.g., Bundy et al. 2009; Robotham et al. 2014) use the estimates of Kitzbichler & White (2008) to calculate $\Gamma(M_*,\, z)$, others (e.g., Conselice et al. 2009; Xu et al. 2012; Conselice 2014) argue that these estimates are incorrect, at least for major mergers. As stated by Hopkins et al. (2010), the dominant source of angular momentum dissipation in major mergers with separations less than $20\, h^{-1}\,$kpc is not dynamical friction, and the merging galaxies are extremely likely by this stage to be in almost radial orbits. These ideas violate the assumptions of Kitzbichler & White (2008), and are further confirmed by analyses of cosmological hydrodynamical simulations combining dark matter and baryonic components (Jiang et al. 2008). Furthermore, L08 presented estimates of τ_m specifically calibrated to the CAS technique of Conselice et al. (2009) using mock galaxy images. Here, as a fiducial case, I only use the estimates of τ_m from L08, specific to estimates of $f_{\rm gm}$ from both galaxy pair counts and CAS analyses. These merger timescales were averaged over both field and cluster environments, and hence account for environmental dependencies. However, the simulation suite of L08 was not large enough to reveal significant mass- or redshift-dependence of τ_m. Hence, I consider it possible that the weak dependencies on these quantities identified by Kitzbichler & White (2008) may be present.

I therefore consider the three following estimates of the times between galaxy mergers:

$$\Gamma(M_*,\, z) = (13.8 \pm 3.1)(1+z)^{-1.6\pm0.6}\,\mathrm{Gyr} \quad (7.6)$$

$$\Gamma(M_*,\, z) = (19.2 \pm 1.5)(1+z)^{-2.2\pm0.2}\,\mathrm{Gyr} \quad (7.7)$$

$$\Gamma(M_*,\, z) = (14.3 \pm 0.6)(1+z)^{-1.6\pm0.6}\,\mathrm{Gyr}, \quad (7.8)$$

based on the work of Conselice et al. (2009), Xu et al. (2012) and Robotham et al. (2014) respectively. While I consider each of Equations 7.6−7.8 to be equally possible, I choose Equation 7.6 (Conselice et al. 2009) as a fiducial prescription. The possible mass- and redshift-dependence of $\Gamma(M_*,\, z)$ is given by the factor $(M_*/10^{10.7}M_\odot)^{-0.3}(1 + z/8)$ (Kitzbichler & White 2008); I further consider it equally likely that this factor is present or absent, while choosing its absence as fiducial. Together, there are then six different possibilities for $\Gamma(M_*,\, z)$ that I consider, each with observational uncertainties. I also demonstrate the effects on the GW signal from binary SMBHs of using systematically larger values of τ_m that are consistent with Kitzbichler & White (2008).

The fitting formulae in Equations 7.6−7.8 are consistent with results at higher redshifts (Conselice 2014). I hence adopt these equations for $z < 3$, and also assume $\frac{dP}{d\log(\mu_*)}$ = constant (Xu et al. 2012). Uncertainties in $\Gamma(M_*,\, z)$ for $z \gtrsim 1$ do not significantly affect my predictions for GW signals from binary SMBHs, because, as I demonstrate, it appears that these signals are dominated by contributions from binary SMBHs at lower redshifts.

7.2.1.2 The Galaxy Stellar Mass Function

I use the latest measurements of the GSMF for $z < 3$ in the range $10^{10}M_\odot \leq M_* \leq 10^{12}M_\odot$ based on the COSMOS/UltraVISTA catalogue (Muzzin et al. 2013). Muzzin et al. (2013) present GSMFs for quiescent (early-type) and star-forming (late-type) galaxies, which were identified using a colour cut. Utilising UV to mid-IR galaxy photometry, with improved sensitivity and sky-coverage over previous compilations, these authors provide the most accurate determinations of the early- and late-type GSMFs currently available.

However, I still need to account for a selection of systematic errors.

Muzzin et al. (2013) use redshift, luminosity and mass measurements obtained through spectral energy distribution analyses. Assuming galaxy magnitude measurements of sufficient accuracy, systematic errors in the photometric redshifts and galaxy stellar mass measurements are dominated by how the stellar populations are modelled (e.g., Bernardi et al. 2010; Mitchell et al. 2013; Courteau et al. 2014). Systematic errors in stellar mass measurements can lead to errors in the GSMF of greater than 0.6 dex (Mitchell et al. 2013). Muzzin et al. (2013) present five separate determinations of the GSMFs of early- and late-type galaxies using different choices for the stellar population synthesis model and star formation history, as well as expanded possibilities for galaxy metallicities and dust attenuation laws. I assume that each of these five GSMF determinations, for which Schechter function fits are given in Table 3 of Muzzin et al. (2013), are equally likely to be correct, but choose the default GSMF of Muzzin et al. (2013), given in their Table 1, as fiducial.

The method of colour selection used to identify early- and late-type galaxies adds further systematic uncertainty to the GSMF estimates. For example, Bernardi et al. (2010) showed that edge-on dusty spiral galaxies are in fact the reddest among the galaxy population, and that more than a third of a red-sequence sample of galaxies could be actively star-forming objects. While Bernardi et al. (2010) suggest further simple morphological selections based on galaxy light concentrations to mitigate these effects, these data were not available in the COSMOS/UltraVISTA catalogue. Instead, I use a crude estimation of the uncertainty range of the GSMF caused by the colour selection from Muzzin et al. (2013), who presented GSMFs determined for significantly different colour cuts to their fiducial scheme (their Table 4). I consider this entire range of variability in the GSMF to be possible. I also demonstrate the effects on the resulting GW signal of possible contamination of colour-selected early-type galaxy samples with late-type galaxies in an extreme scenario by also performing the calculations with the early-type GSMF reduced by 1/3.

7.2.1.3 Relating M_* to $M_{\rm bul}$

I use a robust scheme to relate M_* to $M_{\rm bul}$ for different types of galaxies. Various authors (Simard et al. 2011; Lackner & Gunn 2012; Meert et al. 2014) have measured the ratios of bulge to total light for large galaxy samples from Sloan

Digital Sky Survey (SDSS) data. Based on the work of Simard et al. (2011), Mendel et al. (2014) provide estimates of the bulge to total mass ratios ($M_{\rm bul}/M_*$) for numerous SDSS galaxies, classified as early- or late-type based on colour selections. Emsellem et al. (2011) present a study of the dynamics of a volume-limited sample of early-type galaxies from the ATLAS$^{\rm 3D}$ project, including dynamical mass estimates and accurate morphological identifications.

The scheme I use to relate M_* to $M_{\rm bul}$ for different types of galaxies is summarised as follows.

- Of late-type galaxies with $M_* > 10^{10} M_\odot$, less than 10% have no bulge component (Mendel et al. 2014); in this work, I assume a conservative value of 10%. Of the others, $M_{\rm bul}/M_*$ is in the range 0.2 ± 0.1 (Lackner & Gunn 2012; Mendel et al. 2014; Meert et al. 2014).

- Early-type galaxies with $M_* > 10^{10} M_\odot$ consist of a significant fraction that are best modelled with both bulges and disks, which are identified with the S0 (lenticular) galaxy population (Lackner & Gunn 2012; Mendel et al. 2014; Meert et al. 2014). These galaxies have values of $M_{\rm bul}/M_*$ which are approximately log-normally distributed with mean 0.7 and log-deviation 0.07 dex. A mild correlation between $M_{\rm bul}/M_*$ and M_* may be present for S0 galaxies (Mendel et al. 2014), which I neglect in this work.

- For $10^{10} M_\odot < M_* \lesssim 10^{11.25} M_\odot$, approximately 75% of early-type galaxies are S0s and 25% are true ellipticals (Emsellem et al. 2011). These fractions change to 55% and 45% respectively for larger stellar masses.

While these results are quite approximate, and only derived for a low-redshift ($z \lesssim 0.3$) galaxy sample, I adopt them as a fiducial scheme for relating M_* to $M_{\rm bul}$ for $z < 3$. This scheme is roughly consistent with that used by S13.

In the same way as accounting for scatter in the $M_\bullet - M_{\rm bul}$ relation raises the inferred SMBH mass function (e.g., Aller & Richstone 2002), the bulge mass function inferred from the GSMF will be raised given scatter in relating M_* to $M_{\rm bul}$. Scatter in the $M_{\rm bul} - M_*$ relations can be simply combined with the scatter in the $M_\bullet - M_{\rm bul}$ relation by modifying Equation 7.5 as follows:

$$\frac{dP}{d\log M_\bullet} = \mathcal{N}\left[\log\alpha + \beta\log(M_{\rm bul}(M_*)), \epsilon^2 + \beta^2\sigma_{\rm bul}^2\right], \tag{7.9}$$

where I assume $\sigma_{\rm bul} = 0.1$ for both early- and late-type galaxies. In summary, the function $M_{\rm bul}(M_*)$ in the fiducial model is defined by

$$M_{\rm bul}(M_*) = \begin{cases} 0.2M_*, & \text{for 90\% of late types,} \\ 0.7M_*, & \text{for S0s} \\ M_*, & \text{for ellipticals} \end{cases} \tag{7.10}$$

I demonstrate the effects of possible errors in the fraction of early-type galaxies which are ellipticals by considering cases where this fraction is reduced and increased by 50%.

7.2.1.4 Relating $M_{\rm bul}$ to $M_\bullet$

Despite intense interest in evincing the $M_\bullet - M_{\rm bul}$ relation over the last 15 years, the form of the relation in fact remains uncertain (Kormendy & Ho 2013; Scott et al. 2013). Kormendy & Ho (2013) argue that the $M_\bullet - M_{\rm bul}$ relation is well-modelled by a single power law for all galaxies containing classical bulges which include ellipticals, S0s and spirals with bulges displaying steep central light gradients. However, Scott et al. (2013) find, using an extended version of the galaxy sample of Graham et al. (2011) and independent measurements of $M_{\rm bul}$, that two power laws are required, with a break at $M_{\rm bul} = 3\times10^{10}M_\odot$. A physical distinction between the two power laws was identified by splitting the sample into 'cusp' galaxies with steep power-law central light gradients and galaxies where 'cores', or light-deficits with respect to a cusp, are present. Cusp galaxies are typically of lower masses than core galaxies, and were found by Scott et al. (2013) to have a steeper log-linear $M_\bullet - M_{\rm bul}$ relation than core galaxies.

While I consider the $M_\bullet - M_{\rm bul}$ relations of Kormendy & Ho (2013) and Scott et al. (2013) equally likely, I choose the simpler relation of Kormendy & Ho (2013) as a fiducial case. In Equation 7.9, Kormendy & Ho (2013) find $\alpha = -4.07 \pm 0.05$, $\beta = 1.16 \pm 0.08$ and $\epsilon = 0.29$. Scott et al. (2013) instead find $\alpha = -15.37 \pm 0.18$ and $\beta = 2.22 \pm 0.58$ for $M_{\rm bul} \leq 3\times10^{10}M_\odot$ and $\alpha = -1.86 \pm 0.09$ and $\beta = 0.97 \pm 0.14$ for $M_{\rm bul} > 3\times10^{10}M_\odot$. As Scott et al. (2013) do not estimate the intrinsic scatter, I assume $\epsilon = 0.29$ for the entire range of $M_{\rm bul}$.

I do not consider estimates of the $M_\bullet - M_{\rm bul}$ relation made substantially prior to

Kormendy & Ho (2013) and Scott et al. (2013). Previous estimates are thought to be incorrect because of systematic errors in SMBH and bulge mass estimates, the absence of recently-measured SMBH masses in brightest cluster galaxies, and the presence of galaxies without classical bulges in samples used to fit the relations (see §2.3.1). Various authors infer modest redshift evolution in the $M_\bullet - M_{\rm bul}$ relation such that the typical ratio $M_\bullet/M_{\rm bul}$ may be up to a factor of ~ 3 larger at $z \gtrsim 2$ than the local value (Kormendy & Ho 2013, and references therein). This can be approximately represented by letting $\alpha = \alpha_0 + \log((1+z)^K)$ with $K = 1$ and α_0 as above. As a fiducial case, however, I assume the conservative value of $K = 0$.

7.2.2 GW signals from binary and coalescing SMBHs

In this Chapter, I assume that all binary SMBHs are in circular orbits that evolve only under losses of energy and angular momentum to GWs. While the effects of binary SMBH environments and non-zero orbital eccentricities could modify the GW characteristic strain spectrum from the form in Equation 5.1 at frequencies up to 10^{-8} Hz at the Earth, these effects are highly uncertain. For frequencies $f > 10^{-8}$ Hz within the PTA band (e.g., at $f = f_{\rm yr}$), the characteristic strain spectrum does indeed take the form of Equation 5.1, because the orbits of all binaries radiating GWs at these frequencies are likely to have circularised because of GW-driven evolution. My assumption allows for direct comparison with the majority of other studies on this topic (Jaffe & Backer 2003; Wyithe & Loeb 2003a; Enoki et al. 2004; Sesana et al. 2008b; Kulier et al. 2013; Sesana 2013b), and for the GWB spectrum to be characterised by a single amplitude ($A_{\rm yr}$).

A circular binary SMBH radiates monochromatic GWs at twice its orbital frequency. I use Equation 3.34 to define the rms GW strain amplitude, h_s, radiated by a circular binary, and use an expression for the rms GW-induced sinusoidal variations to the pulse times of arrival (ToAs) from radio pulsars, σ_R, from Equation 20 of Sesana et al. (2009):

$$\sigma_R = \frac{8}{15}\sqrt{\frac{5}{32}} h_s \left(\frac{T}{f}\right)^{1/2}. \tag{7.11}$$

Both h_s and σ_R are averaged over all binary orientation parameters. The strain amplitude, $h_{\rm mem}$, of a memory burst from a coalescing binary SMBH is defined in Equation 4.8.

To calculate the GWB amplitude, consider a multivariate density function, $f_{\mathbf{X}}$, for the observed binary SMBH coalescence rate, R, in terms of a k-component parameter vector $\mathbf{X}$ with components X_i indexed by an integer i:

$$f_{\mathbf{X}} = \prod_{i=1}^{k} \frac{\partial[R]}{\partial X_i}. \tag{7.12}$$

Then, $A_{\rm yr}$ is given by

$$A_{\rm yr} = \left[f_{\rm yr} \int ... \int_{\mathbf{X}} f_{\mathbf{X}} \left(\frac{dt}{df} h_s^2 \right)_{f=f_{\rm yr}} dX_1 ... dX_k \right]^{1/2}. \tag{7.13}$$

Here, $\frac{dt}{df} = \left(\frac{df}{dt}\right)^{-1}$ for the domains of t and f under consideration.

7.2.3 Assembling the model

Mergers between galaxies containing bulges with masses M_* and $M_*\mu_*$ come in nine types, because the galaxies with each mass may be either elliptical, S0 or late-type. In each case, a different prescription is required to identify the bulge masses of the merging galaxies, and hence the masses of the SMBHs in the merging galaxies. Consider a merger between a galaxy of type i, with mass M_*, and a galaxy of type j, with mass $M_*\mu_*$, where i and j each denote either an elliptical, S0 or late-type galaxy. The fraction of cases where this merger will occur is given by $\frac{\Phi_{*,\,j}}{\Phi_*}$, where Φ_* is given by Equation 7.4 and the mass functions are evaluated at a mass $M_*\mu_*$. For early-type galaxies, $\Phi_{*,\,j}$ is specified according to the fractions of ellipticals and S0s at different masses, and for late-type galaxies, $\Phi_{*,\,j}$ is simply the fraction which contain bulges. The SMBH masses corresponding to the galaxies of types i and j, $M_{\bullet,\,i}$ and $M_{\bullet,\,j}$ respectively, are described by the probability density function in Equation 7.9 for $M_{\rm bul}$ given by $M_{{\rm bul},\,i}(M_*)$ and $M_{{\rm bul},\,j}(M_*\mu_*)$ respectively. Hence,

in order to calculate $A_{\rm yr}$, I combine Equations 7.2, 7.3, 7.4, 7.9 and 7.13 as follows:

$$\begin{aligned}\frac{A_{\rm yr}^2}{f_{\rm yr}} = &\int_{\log(10^{10}M_\odot)}^{\log(10^{12}M_\odot)} d\log M_* \int_0^3 dz \int_{\log(1/3)}^{0} d\log\mu_* \\ &\times \int_{-\infty}^{\infty} d\log M_{\bullet,\,i} \int_{-\infty}^{\infty} d\log M_{\bullet,\,j} \\ &\times \frac{4\pi d^2 V_c}{d\Omega dz}\frac{1}{\Gamma}\frac{dP}{d\log(\mu_*)}\frac{dt_p}{dt}\left(\frac{dt}{df}\right)_{f=f_{yr}} \\ &\times \frac{dP}{d\log M_{\bullet,\,i}}\Big|_{M_*}\frac{dP}{d\log M_{\bullet,\,j}}\Big|_{M_*\mu_*} \\ &\times \sum_{i,\,j}\Phi_{*,\,i}\frac{\Phi_{*,\,j}}{\Phi_*}h_s^2(M_{\bullet,\,i},\,M_{\bullet,\,j},\,z,\,f_{\rm yr}).\end{aligned} \tag{7.14}$$

I evaluate this integral numerically by summing over the integrand in bins of $\log M_*$, z, $\log\mu_*$, $\log M_{\bullet,\,i}$, and $\log M_{\bullet,\,j}$. To determine the predicted numbers of CW sources, I count the numbers of individual binary SMBHs in each bin with different values of h_s radiating GWs at $f_{\rm yr}$ within a nominal bandwidth of $\Delta f = (10\,{\rm yr})^{-1}$. I also record the rate of memory bursts in each bin with corresponding amplitudes $h_{\rm mem}$. These latter operations are equivalent to numerically evaluating the conditional densities of GW sources in terms of h_s and $h_{\rm mem}$.

Equation 7.14 builds on the approach of S13 in two ways. First, I account for the effects of intrinsic scatter in the $M_\bullet - M_{\rm bul}$ relation and in relating M_* to $M_{\rm bul}$. I also attempt to match the numbers of galaxy mergers of different types to the measured GSMFs, rather than assuming the same galaxy pair fractions for all types of mergers.

7.2.4 Summary of assumptions

The model presented above for the binary SMBH population emitting GWs in the PTA band includes observational estimates of the average time between galaxy mergers, the $M_\bullet - M_{\rm bul}$ relation and the GSMFs of early- and late-type galaxies. The model also includes a scheme to relate $M_{\rm bul}$ to M_*. The key assumptions in this model are as follows.

1. The early- and late-type GSMFs, the average times between major ($\mu_* \geq 1/3$) mergers and the probabilities of different values of μ_*, for massive galaxies with $M_* > 10^{10}M_\odot$ and $z < 3$, are accurately described by the current observations.

2. The relative numbers of massive major mergers corresponding to the nine galaxy type combinations I consider are determined only by the GSMFs.

3. Every massive galaxy contains a central SMBH with a mass that can be inferred using empirical $M_{\rm bul} - M_*$ and $M_\bullet - M_{\rm bul}$ relations.

4. In each galaxy merger remnant, the central SMBHs form a bound binary that circularises before an observed GW emission frequency at the Earth of $f_{\rm yr}$ is reached.

5. Binary SMBH coalescence in merger remnants occurs within times that are much shorter than the times between major mergers.

7.3 Results

7.3.1 The GWB amplitude

I first calculated $A_{\rm yr}$ using Equation 7.14 given the fiducial prescriptions for $\Gamma(M_*, z)$, the GSMF, the scheme relating M_* and $M_{\rm bul}$ and the $M_\bullet - M_{\rm bul}$ relation, as detailed in §7.2.1. The resulting fiducial value for $A_{\rm yr}$ was 1.3×10^{-15}. I then identified the possible ranges of $A_{\rm yr}$ consistent with the observational uncertainties in each of $\Gamma(M_*, z)$, the GSMF and the $M_\bullet - M_{\rm bul}$ relation alone. This was accomplished by generating 600 realisations of $A_{\rm yr}$ with the parameters of a single one of these quantities randomised and with the other terms in Equation 7.14 held fixed at their fiducial values. This process was computationally expensive, as each realisation required three minutes to complete on 8 CPU cores. The 600 realisations were sufficient to capture the range of variability in the observational uncertainties, as all posterior predictive parameter distributions derived from observations were assumed to be normal. The process was then repeated with randomisation individually in the other two quantities.

For $\Gamma(M_*, z)$, I considered six possible expressions equally likely and also drew randomly from the parameter uncertainty ranges in each expression. For the GSMF, I considered five possible sets of Schechter function fits corresponding to different galaxy stellar mass measurement techniques and the observational uncertainties of each, as well as variations given uncertainty in the colour selection of early- and late-type galaxies. Finally, for the $M_\bullet - M_{\rm bul}$ relation, I considered two possible

relations and their associated parameter uncertainties. I did not consider any 'observational' uncertainties in the scheme for relating galaxy stellar masses to bulge masses, including the respective fractions of elliptical, S0 and late-type disk+bulge galaxies.

Histograms of the resulting three samples of realisations of $A_{\rm yr}$ are shown in the middle three panels of Figure 7.1, along with the fiducial value of $A_{\rm yr}$ (as a vertical dashed line). While the possible ranges of $A_{\rm yr}$ given observational uncertainties in $\Gamma(M_*, z)$ and the GSMF are roughly equivalent, observational uncertainty in the $M_\bullet - M_{\rm bul}$ relation results in a slightly larger range of possible $A_{\rm yr}$ values.

I also considered the effects of adopting four modifications to the fiducial model relating to parameters for which current observational constraints are poor. These modifications, which were introduced in §7.2.1, result in the following values of $A_{\rm yr}$:

1. When I decrease the early-type GSMF by a factor of 1/3 to simulate an extreme case of contamination of colour-selected early-type galaxy samples by late-type galaxies (e.g., edge-on spirals), I obtain $A_{\rm yr} = 10^{-15}$. This represents a decrease of 0.12 dex over the fiducial model.

2. When I adopt the massive galaxy merger timescale from Kitzbichler & White (2008) with the associated mass- and redshift-dependence, rather than from numerical simulations of galaxy mergers (L08), I obtain $A_{\rm yr} = 7.4 \times 10^{-16}$. This represents a decrease of 0.24 dex over the fiducial model.

3. When I introduce a redshift-dependent normalisation, α, of the $M_\bullet - M_{\rm bul}$ relation with $K = 1$ such that the normalisation is a factor of three greater at $z = 2$, I obtain $A_{\rm yr} = 1.8 \times 10^{-15}$. This is an increase of 0.14 dex over the fiducial model.

4. Finally, in exploring the effects of either reducing or increasing the fraction of early-type galaxies which are ellipticals by 50%, I obtain an associated variation in $A_{\rm yr}$ of 5% (0.02 dex).

The first three modifications are clearly significant, as compared to the fourth: I depict the resulting values of $A_{\rm yr}$ in the top panel of Figure 7.1. While the effects of modifications (1) and (3) are comparable in magnitude, adopting the galaxy merger

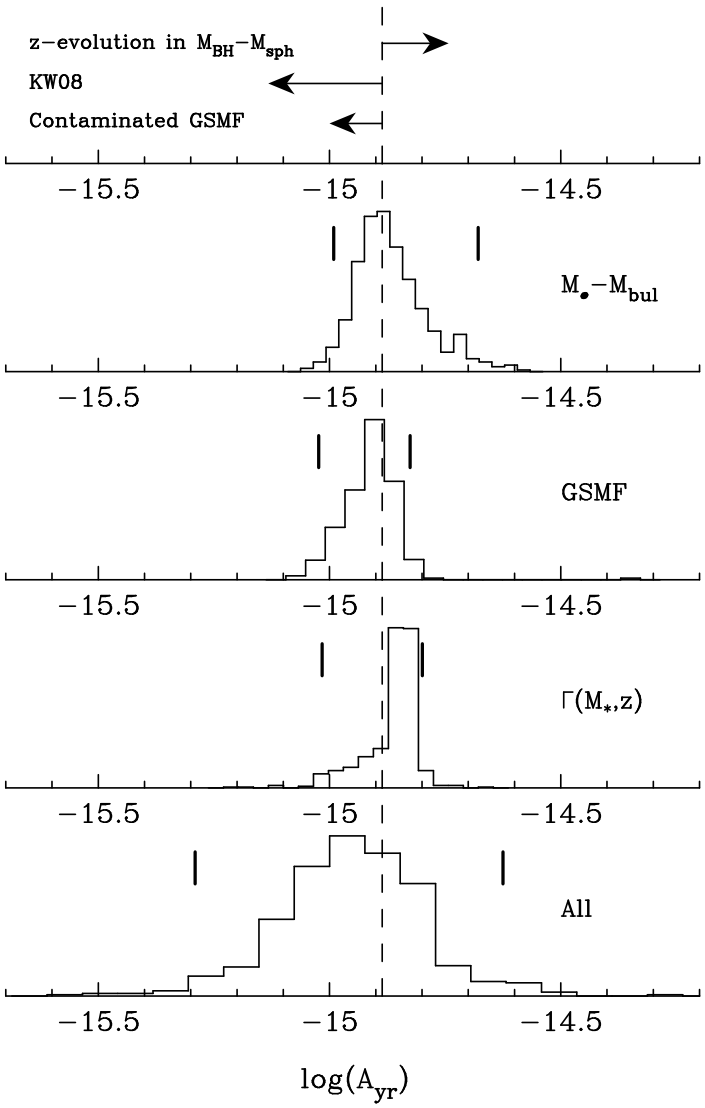

Figure 7.1: Depiction of uncertainties in the value of $A_{\rm yr}$ calculated using Equation 7.14. The standardised histograms labelled '$\Gamma(M_*, z)$', 'GSMF' and '$M_\bullet - M_{\rm bul}$' show the distributions of 600 realisations of $A_{\rm yr}$ given randomisation over the prescriptions for the respective quantities alone. The vertical dashed line indicates the value of $A_{\rm yr} = 1.3 \times 10^{-15}$ resulting from the fiducial prescriptions for all quantities in Equation 7.14. The three arrows at the top of the Figure show how much this fiducial value varies given possible systematic uncertainties in the model. From the bottom, the arrowheads indicate the values of $A_{\rm yr}$ corresponding to a possibly contaminated early-type GSMF, galaxy merger timescales consistent with Kitzbichler & White (2008) and redshift-evolution in the normalisation of the $M_\bullet - M_{\rm bul}$ relation respectively (see text for details). The standardised histogram labelled 'All' shows the distribution of 600 realisations of $A_{\rm yr}$ given randomisations over all uncertainties considered in this work. For all four histograms, the 2.5% and 97.5% percentiles are shown as thick vertical bars.

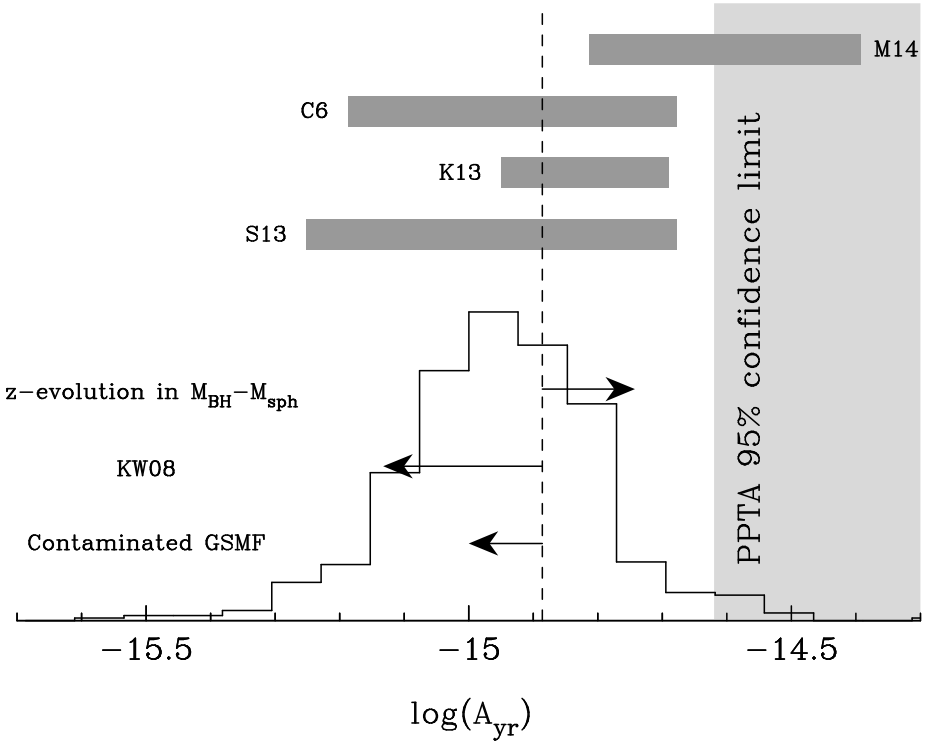

Figure 7.2: Comparison between the predictions for $A_{\rm yr}$ and those from Chapter 6 and other studies. The histogram is identical to that labelled 'All' in Figure 7.1. The vertical dashed line indicates the fiducial prediction of $A_{\rm yr} = 1.3 \times 10^{-15}$, and the arrows indicate systematic uncertainties in $A_{\rm yr}$ (see the caption of Figure 7.1 for details). The dark grey horizontal bars show 68% confidence intervals for $A_{\rm yr}$ predicted by S13, Kulier et al. (2013), labelled 'K13', Chapter 6, labelled 'C6', and McWilliams et al. (2014), labelled 'M14'. The light grey shaded area indicates the 95% confidence PPTA upper limit on the GWB amplitude, set at $A_{\rm yr} = 2.4 \times 10^{-15}$.

timescales of Kitzbichler & White (2008) makes a large difference to the prediction of $A_{\rm yr}$. This is expected, because the Kitzbichler & White (2008) merger timescales are roughly a factor of three longer than those of L08. Indeed, modification (2) results in a value of $A_{\rm yr}$ that is lower than the 2.5% percentile of the distributions of $A_{\rm yr}$ values given the three observational uncertainties considered so far.

As an illustration of the full range of possible values of the GWB amplitude given the uncertainties considered in $\Gamma(M_*, z)$, the GSMF and the $M_\bullet - M_{\rm bul}$ relation combined with modifications (1)−(3) listed above, I generated a new sample of 600 realisations of $A_{\rm yr}$. In this case, I simultaneously randomised over $\Gamma(M_*, z)$, the GSMF and the $M_\bullet - M_{\rm bul}$ relation as described above, and also (*i*) decreased the early-type GSMF by a factor uniformly drawn from the interval $[0, 1/3]$, (*ii*) set the galaxy merger timescale at a value uniformly drawn between the predictions of L08 and Kitzbichler & White (2008) (neglecting any mass- or redshift- dependence), and (*iii*) set the redshift-evolution index K of the normalisation of the $M_\bullet - M_{\rm bul}$ relation to a number uniformly drawn from the interval $[0, 1]$. A histogram of the resulting sample of realisations of $A_{\rm yr}$ is shown in the bottom panel of Figure 7.1. The 95% confidence interval on $A_{\rm yr}$, considering all uncertainties, is $5.1 \times 10^{-16} < A_{\rm yr} < 2.4 \times 10^{-15}$.[1]

I next compare the results for $A_{\rm yr}$ with earlier predictions. In Figure 7.2, I again show the histogram of realisations of $A_{\rm yr}$ corresponding to randomisation over all uncertainties, as well as the values of $A_{\rm yr}$ corresponding to modifications (1)−(3) listed above. Above these, I show the 68% confidence intervals on $A_{\rm yr}$ from S13, Chapter 6, and the recently-published models of Kulier et al. (2013) and McWilliams et al. (2014). The predictions that I consider all account for the most recent determinations of the $M_\bullet - M_{\rm bul}$ relation (Kormendy & Ho 2013; Scott et al. 2013).

The range of possible values of $A_{\rm yr}$ predicted by S13 is consistent with (albeit somewhat broader than) the range I predict given all uncertainties that I consider in here. Both the present work and S13 attempt to synthesise all uncertainties in

[1] One unaccounted-for uncertainty is in the adopted cosmological parameters. In this Chapter, I choose to use the latest results from the *Planck* satellite (Planck Collaboration et al. 2013), whereas in previous Chapters I used the WMAP1 results (Spergel et al. 2003). Despite the substantial differences in the values of H_0, Ω_Λ and Ω_M between these two sets of parameters, the fiducial value of $A_{\rm yr}$ decreases by only 2% when the *Planck* parameters are replaced by the WMAP1 parameters. Hence, uncertainty in cosmological parameters is unlikely to significantly affect the 95% confidence interval on $A_{\rm yr}$.

quantities relevant to characterising the SMBH-SMBH coalescence rate, and use the same underlying model assumptions to predict the GWB amplitude. My range of predictions is less extended than that of S13 because of the greater uncertainty assumed by S13 in the GSMF and the galaxy merger rate.[2]

A semi-analytic approach (Guo et al. 2011) was used in Chapter 6 to predict SMBH-SMBH coalescence rates within the Millennium simulation (Springel et al. 2005), coupled with prescriptions for binary SMBH orbital evolution in stellar environments (Sesana 2010). I found that the characteristic strain spectrum may be attenuated relative to the case of circular binary orbits and GW-driven evolution at frequencies $f \lesssim 10^{-8}\,\mathrm{Hz}$. However, the 68% confidence interval on the characteristic strain spectral amplitude at a frequency of $f_{\rm yr}$ (see, e.g., Figure 6.4) is consistent with the range of values of $A_{\rm yr}$ I find in this Chapter.

The prediction of Kulier et al. (2013) is derived from hydrodynamic numerical galaxy formation simulations in cluster and field environments, but may be biased relative to semi-analytic galaxy formation models implemented in large-volume numerical dark matter simulations because of the specific choice of overdense and underdense regions to study. However, the prediction of Kulier et al. (2013) naturally includes a particularly sophisticated treatment of galaxy merger timescales.

McWilliams et al. (2014) suggest a model for the binary SMBH population which includes the assumption that all evolution in the early-type GSMF at $z < 1$ is driven by galaxy mergers; however, their predicted GWB amplitude is larger than current PTA constraints (Shannon et al. 2013). This model would necessarily include a shorter galaxy merger timescale than that predicted by L08 in order to maintain consistency with the observed numbers of merging galaxies. Overall, besides the study of McWilliams et al. (2014), it is encouraging that different models appear to agree on the amplitude of the characteristic strain spectrum from binary SMBHs. In particular, the upper ends of most predicted ranges of $A_{\rm yr}$ all appear to be consistent.

In Figure 7.2, I also depict the best existing 95% confidence upper limit on $A_{\rm yr}$ from Shannon et al. (2013) as a shaded region. Some realisations of $A_{\rm yr}$ given observational uncertainties in the model are inconsistent with this upper limit. However, the upper limit is generally consistent with my model given all uncertainties.

[2]Some methodological differences also exist between the present work and S13 in how different realisations of $A_{\rm yr}$ were obtained.

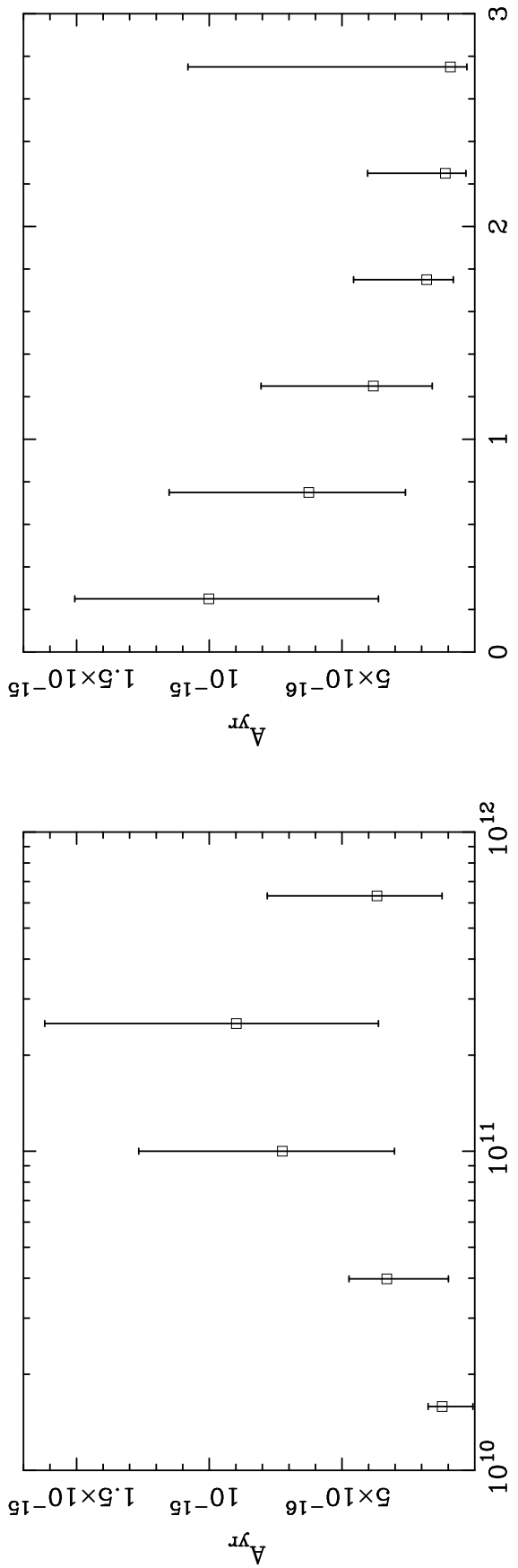

Figure 7.3: Left: Values of $A_{\rm yr}$ from binary SMBHs created in major mergers involving galaxies of different stellar masses. The squares indicate the fiducial model result, and the vertical error bars indicate 95% confidence intervals from 600 realisations of the model with all uncertainties that I account for. *Right:* Values of $A_{\rm yr}$ from binary SMBHs in six redshift bins in the interval $0 < z < 3$. The squares and error bars are as in the left panel. The redshift intervals correspond to the ranges within which the GSMF was evaluated by Muzzin et al. (2013).

In Figure 7.3, I plot the values of $A_{\rm yr}$ predicted by the fiducial model in different ranges of M_* (left panel) and z (right panel). I also show the 95% confidence intervals on these values given all uncertainties I consider. The galaxy mass ranges correspond to the values of M_* of the larger galaxies in mergers. The dominant contributions to the GWB are from binary SMBHs formed in mergers involving galaxies with $M_* \gtrsim 5 \times 10^{10} M_\odot$, and from binary SMBHs at redshifts $z \lesssim 1.5$. The confidence intervals that I provide further suggest that contributions to the GWB from outside these ranges are not significant.[3] Finally, binary SMBHs created in mergers involving at least one late-type galaxy correspond to $A_{\rm yr} = 4.7 \times 10^{-16}$, whereas mergers involving only early-type galaxies correspond to $A_{\rm yr} = 1.2 \times 10^{-15}$. Hence, within the present model, the GWB is likely to be dominated by galaxy mergers involving only early-type galaxies (S0s and ellipticals).

7.3.2 Individual GW sources: continuous waves and memory bursts

In the process of evaluating Equation 7.14, I also calculated the numbers of individual binary SMBHs that produce monochromatic (CW) GW signals, along with the numbers of GW memory bursts emitted during SMBH-SMBH coalescence events. I integrated over the counts of individual binaries emitting GWs at frequencies $f = f_{\rm yr}$ in a frequency bin of width $\Delta f = (10\,{\rm yr})^{-1}$ and evaluated the numbers of binaries with different GW strain amplitudes h_s. These results are shown in the left panel of Figure 7.4 for the fiducial model as well as for two variations to the fiducial model (modifications (2) and (3) listed above). The smoothness of the curves in Figure 7.4 reflects the analytic nature of the calculation. I also show results for the fiducial model while restricting the source counts to binaries at redshifts $z < 1$ and with the more massive progenitor galaxy mass $M_* > 10^{11} M_\odot$. The restricted source counts are identical to the full source counts for $h_s \gtrsim 2 \times 10^{-15}$. From Equation 7.11, the characteristic amplitude of the sinusoidal ToA variations induced by a binary SMBH with strain amplitude h_s at $f = f_{\rm yr}$, over a 10 yr observation, is $\sigma_R = 21 (h_s / 10^{-15})\,{\rm ns}$.

[3] While the most massive galaxies do not appear to contribute significantly to the GWB, it is apparent from, e.g., Figure 6 of Muzzin et al. (2013) (see also Baldry et al. 2012) that the Schechter function fits to the early-type GSMFs under-predict the observed GSMF at masses $M_* \gtrsim 10^{11.5}$.

Figure 7.4: *Left:* the counts of individual sources at and above given GW strain amplitudes (h_s) at a GW frequency of $f_{\rm yr}$ in a frequency bin of width $\Delta f = (10\,{\rm yr})^{-1}$. *Right:* the numbers of GW memory bursts per year at and above given strain amplitudes $h_{\rm mem}$. In both panels, the results of the fiducial model are shown as thick black solid curves, the results from a model with maximal redshift evolution in the $M_\bullet - M_{\rm bul}$ relation ($K = 1$ corresponding to α in Equations 7.5 and 7.9 increased by a factor of three at $z = 2$) are shown as dotted red curves, and the results from a model with galaxy merger timescales consistent with Kitzbichler & White (2008) are shown as blue dashed curves. The green thin solid curves represent source counts for the fiducial model evaluated with the restrictions $z < 1$ and $M_* > 10^{11} M_\odot$.

Scaling these CW source counts to other GW frequencies is non-trivial. The GW strain amplitude of a binary SMBH radiating at a frequency f can be expressed as $h_s = h_{s,\,\mathrm{yr}}(f/f_{\mathrm{yr}})^{2/3}$, where $h_{s,\,\mathrm{yr}}$ is the strain amplitude radiated by that binary at a frequency f_{yr}. Furthermore, the total number of binaries per unit frequency radiating GWs at a frequency f is related to the number of binaries per unit frequency radiating GWs at f_{yr} by the factor $(f/f_{\mathrm{yr}})^{-11/3}$, assuming GW-driven binary orbital evolution. Then, the number of binaries per unit frequency emitting GWs at or above a strain amplitude of h_s, at a frequency f, may be written as $n(f, h_s) = n(f_{\mathrm{yr}}, h_s(f/f_{\mathrm{yr}})^{-2/3})(f/f_{\mathrm{yr}})^{-11/3}$. For example, while the fiducial model predicts $\sim 10^{-2}$ CW sources with $h_s \geq 10^{-15}$ in a frequency bin of width $\Delta f = (10\,\mathrm{yr})^{-1}$ at $f = f_{\mathrm{yr}}$, this prediction changes to ~ 0.1 sources at $f = f_{\mathrm{yr}}/5$ with $h_s \geq 10^{-15}$ in the same frequency bin width.

I can hence directly compare my predicted CW source counts with the work of Sesana et al. (2009). These authors considered a wide variety of SMBH growth scenarios within the framework of a semi-analytic model for galaxy formation (Bertone et al. 2007) implemented in the Millennium simulation results (Springel et al. 2005). I directly compare predictions for the number of binary SMBHs inducing ToA variations with characteristic amplitudes $\sigma_R \geq 30\,\mathrm{ns}$. For consistency, I consider an observation time span of $T = 5\,\mathrm{yr}$ and GW frequencies $f > 3 \times 10^{-9}\,\mathrm{Hz}$, and integrate over the number of sources per unit frequency with $\sigma_R \geq 30\,\mathrm{ns}$ in the range $3 \times 10^{-9} - 10^{-7}\,\mathrm{Hz}$ (integrating to higher frequencies does not significantly alter my results). I neglect the issue of whether these signals are resolvable given the presence of a GWB. I predict 0.6 CW sources with $\sigma_R \geq 30\,\mathrm{ns}$ for the fiducial model, 0.1 CW sources for a pessimistic model assuming the galaxy merger timescales of Kitzbichler & White (2008), and 1.2 CW sources for the optimistic model with significant redshift-evolution in the $M_\bullet - M_{\mathrm{bul}}$ relation. Sesana et al. (2009) predict between 0.05 and 3 such sources (their Figure 3), which is consistent with my results.

I also predicted the numbers of binary SMBH coalescence events per observed year at or above a given GW memory burst amplitude, h_{mem} (see Equation 4.8) for $h_{\mathrm{mem}} > 10^{-16}$. The results are shown in the right panel of Figure 7.4, again for the fiducial model and two variations to this model. I also again show results for the fiducial model with the restrictions of $z < 1$ and $M_* > 10^{11} M_\odot$; for $h_{\mathrm{mem}} \gtrsim 6 \times 10^{-16}$, the restrictions make no significant difference.

In summary, the expected numbers of individual GW sources predicted by my empirical binary SMBH model are small. At most ~ 1 CW source is expected to induce ToA variations with characteristic amplitudes $\geq 30\,\mathrm{ns}$ over a 5 yr observation time span. Also, approximately one GW memory burst with $h_{\mathrm{mem}} > 5 \times 10^{-16}$ is expected every 1000 yr.

7.4 Implications for GW detection with PTAs

7.4.1 The GWB from binary SMBHs

The future sensitivities of PTAs to the GWB are the subjects of ongoing research (e.g., Siemens et al. 2013; Moore et al. 2014; Hobbs et al. 2014). For example, future pulsar observing systems and cadences, new pulsar discoveries, the effects of the interstellar medium and pulsar timing noise characteristics, all of which significantly affect PTA sensitivities, are difficult to forecast because of a lack of quantitative, predictive models. An idealised treatment of the problem by Siemens et al. (2013) suggests that, for the NANOGrav collaboration, a GWB with amplitude $A_{\mathrm{yr}} = 10^{-15}$ may be detectable by the year 2021. I note that Siemens et al. (2013) assumed that the GWB characteristic strain spectrum has the power law form given in Equation 5.1. I find in this Chapter that the GWB amplitude is likely to be in the range $5.1 \times 10^{-16} < A_{\mathrm{yr}} < 2.4 \times 10^{-15}$ with 95% confidence. If the GWB amplitude were to lie in the upper part of this range, as is expected given the more commonly preferred major galaxy merger timescale (L08), I suggest that detecting a GWB from binary SMBHs is indeed an attainable, short-term goal for PTAs.[4]

What can PTA upper limits on or detections of the GWB reveal about the determinants of the GWB amplitude? The GWB may be parameterised by a single number, A_{yr}, at least at GW frequencies $f \gtrsim 10^{-8}\,\mathrm{Hz}$, the value of which is dependent on myriad quantities. Useful information can be gleaned if one of these quantities is particularly unconstrained otherwise. For example, if one remains agnostic with respect to the galaxy merger timescale, a particular value of this timescale would correspond to a range of possible GWB amplitudes given knowledge of all the other determinants of A_{yr}. Then, a PTA constraint on A_{yr} would correspond to a con-

[4] If I calculate the range of possible A_{yr} values given all uncertainties, while assuming the L08 galaxy merger timescales, I find $A_{\mathrm{yr}} > 9 \times 10^{-16}$ with the 97.5% confidence.

straint on the galaxy merger timescale, given the assumptions inherent in my model. Through such exercises, PTAs could directly impact understanding of galaxy and SMBH growth, in a more general sense than by testing specific GWB models using PTA data. I leave a demonstration of such techniques for Chapter 8.

It is worth noting that a significantly more sensitive PTA data set is, however, required to detect a GWB of a given amplitude than to exclude a GWB of that amplitude. Pulsar timing data are affected by many noise processes, some of which cause correlations between timing measurements of different pulsars (e.g., Foster & Backer 1990; Cordes & Shannon 2010). Unlike the procedure of forming upper limits on $A_{\rm yr}$, the detection of a GWB requires that its effects on a PTA data set must be shown to be isolated from all noise processes. In practice, this relies on the pairwise correlations between multiple contemporaneous pulsar timing data sets being shown to be consistent with the Hellings & Downs function of the angular separation between pulsars on the sky (Hellings & Downs 1983).

7.4.2 CW signals from individual binary SMBHs

Future PTA observations with planned telescopes such as the Five Hundred Metre Aperture Spherical Telescope (FAST, Li, Nan & Pan 2013) and the Square Kilometre Array (SKA, Cordes et al. 2004) may include up to 100 pulsars with timing noise standard deviations of $\sim$100 ns (Lazio 2013; Hobbs et al. 2014). Ellis et al. (2012) constructed theoretical PTA sensitivity curves using simulated data sets with both 100 arbitrarily-located pulsars or 17 pulsars at the locations of the best-timed pulsars observed by the NANOGrav collaboration, in all cases with timing noise standard deviations of 100 ns and 5 yr observation times. These sensitivity curves, shown in Figure 4.5, represent the values of h_s at different frequencies at which the probability of a false detection was less than 10^{-4} in 95% of realisations of their simulated data sets. Importantly, the sensitivity curves were averaged over all source positions and orientations, and account for pulsar parameter fitting. I predict the numbers of detectable sources for PTAs with these sensitivity curves by evaluating the following integral:

$$N_{\rm detect} = \int_{(10\,{\rm yr})^{-1}}^{10^{-7}\,{\rm Hz}} \frac{dF[h_{\rm sens}(f)]}{df} df, \tag{7.15}$$

where $h_{\rm sens}(f)$ is the sensitivity curve and $\frac{dF(h_{\rm sens})}{df}$ is the predicted number of sources with strain amplitudes $h_s \geq h_{\rm sens}(f)$ per unit frequency at a frequency f. The sensitivities of PTAs to CW sources are generally poor for frequencies $f \gtrsim 10^{-7}\,{\rm Hz}$ and few sources are expected at these frequencies.

Using my predictions for the numbers of CW sources, I evaluate $\frac{dF(h_{\rm sens})}{df}$ by scaling the predictions as described in §7.3.2. Then, for the fiducial model and for the two sensitivity curves of Ellis et al. (2012) corresponding to their coherent $\mathcal{F}$-statistic, I obtain predictions of 0.07 and 1.3 detectable sources for the 17- and 100-pulsar cases respectively. For the restricted fiducial model, corresponding only to sources with $z < 1$ and $M_* > 10^{11}\,M_\odot$ these reduce marginally to 0.06 and 1 source respectively. For the optimistic case with strong redshift-evolution of the $M_\bullet - M_{\rm bul}$ relation, I obtain predictions of 0.2 and 2.8 detectable sources for the 17- and 100-pulsar cases respectively. In contrast, the current PPTA sensitivity curve produced by Zhu et al. (2014) corresponds to $\lesssim 10^{-4}$ detectable sources. 'Noise' caused by the summed GW signal from the binary SMBH population will further increase the difficulty of detecting individual binaries (see Chapter 5 and Sesana et al. 2009).

7.4.3 GW memory bursts from coalescing binary SMBHs

A PTA data set with 20 pulsars timed with a precision of 100 ns for 10 yr is sensitive to memory bursts with amplitudes $h_{\rm mem} > 5 \times 10^{-15}$ over $70 - 80\%$ of the data span (van Haasteren & Levin 2010; Cordes & Jenet 2012). As the sensitivity of such an idealised PTA to memory bursts scales roughly as the square root of the number of pulsars (van Haasteren & Levin 2010), a PTA with 100 pulsars timed with 100 ns precision for 10 yr may be sensitive to memory bursts with $h_{\rm mem} > 2\times10^{-15}$. However, my results suggest that only $\sim10^{-5}$ bursts with $h_{\rm mem} > 5\times10^{-15}$ and $\sim10^{-3}$ bursts with $h_{\rm mem} > 2 \times 10^{-15}$ are expected over 10 yr. Thus, under the model presented here, GW memory bursts from coalescing binary SMBHs do not represent viable sources for PTAs.

7.5 Discussion

The predictions in this Chapter for the GWB amplitude, $A_{\rm yr}$, are conservative within their respective scenarios, for a number of reasons. (i) I do not account

for minor galaxy mergers with stellar mass ratios $\mu_* < 1/3$, or for mergers where the more massive galaxy has a mass $M_* < 10^{10} M_\odot$. (*ii*) I do not consider the possibility of gas accretion onto SMBHs prior to coalescence during galaxy mergers (e.g., Van Wassenhove et al. 2012), which would raise the SMBH masses and hence the emitted GW amplitudes (e.g., Sesana et al. 2008b). (*iii*) The most massive galaxies are typically found in cluster environments, where times between galaxy mergers may be shorter (cf. Lotz et al. 2013), implying a higher merger rate for these galaxies and hence a higher GW signal. However, I do not expect the inclusion of these factors to significantly affect the predicted GWB amplitudes. I reiterate that the effects of interactions between binary SMBHs and their environments are unlikely to affect the predictions for the GWB amplitude at frequencies $f \gtrsim 10^{-8}\,\mathrm{Hz}$, such as at $f_{\rm yr}$. This is because the orbital evolution of binary SMBHs radiating GWs at these frequencies is expected to be predominantly GW-driven, which further leads to the circularisation of the orbits.

Of all sources of uncertainty I consider in predicting the GWB amplitude given relevant observational quantities, the choice of galaxy merger timescale dominates the range of possible GWB amplitudes. Furthermore, the merger timescale may be even more uncertain than the range spanned by the predictions I consider (L08; Kitzbichler & White 2008). The simulations of L08 were conducted only for mergers between gas-rich disk galaxies, some of which contained small bulges, whereas I find that the GWB is likely dominated by binary SMBHs formed in mergers solely between early-type galaxies. Further theoretical studies of galaxy merger timescales for early-type systems are clearly required in order to better predict the GWB amplitude. The dominance of low-redshift ($z \lesssim 1.5$) early-type major mergers of massive ($M_* \gtrsim 5 \times 10^{10} M_\odot$) galaxies in determining the GWB amplitude is a further important consequence of this work for both theoretical and observational studies of galaxy mergers aimed at informing PTA research.

The other significant source of uncertainty in my predictions is in the $M_\bullet - M_{\rm bul}$ relation, both in its local form and in its possible redshift-evolution. In contrast to uncertainty in the galaxy merger timescale, it is likely that this uncertainty will only be resolved through further observations which significantly expand the sample of known SMBH masses. Promisingly, Davis et al. (2013) report that hundreds of SMBH mass measurements may be possible with the Atacama Large Millimetre Array (ALMA).

Under what circumstances could the GWB amplitude lie outside the range I predict given all uncertainties that I consider? The predicted range of GWB amplitudes, $5.1 \times 10^{-16} < A_{\rm yr} < 2.4 \times 10^{-15}$, encompasses all purely observational uncertainties, as well as uncertainty ranges that I set for other quantities for which observational constraints are poor, such as the galaxy merger timescale. It may be possible that these latter ranges are incorrect. Furthermore, not all galaxies may host a central SMBH, as I have assumed. The interaction between a binary SMBH and a third SMBH would likely cause the least massive SMBH to be ejected (e.g., Gerosa & Sesana 2014), lowering the number of coalescing SMBHs. If not every massive galaxy at $z \sim 1$ formed with a central SMBH, the GWB amplitude would again be lowered. It may also be possible that binary SMBHs do not always coalesce on timescales less than the times between galaxy mergers.

The presence of a few strong GW emitters among the binary SMBH population implies that some excess, non-Gaussian scatter will be present in the GW signals produced by this population. The magnitude of this excess scatter in $A_{\rm yr}$ depends on exactly how many binary SMBH systems contribute significantly to the GWB. I suggested in Chapter 5 that the statistics of ToA variations induced by GWs from binary SMBHs are mildly non-Gaussian for frequencies $f > f_{\rm yr}/5$ because of appreciable contributions to the squared characteristic strain spectrum, $h_c^2(f)$, from individual binaries at every GW frequency. Figure 7.5 shows the number of binary SMBHs in the fiducial model of this Chapter corresponding to galaxy mergers with primary stellar masses greater than or equal to a given M_* (top), as well as the fractions of $A_{\rm yr}^2$ contributed by these binaries (bottom). I show in particular binaries radiating at a GW frequency of $f_{\rm yr}$ in a frequency bin of width $\Delta f = (10\,{\rm yr})^{-1}$. The fiducial model suggests that the contributions of individual GW sources to $h_c^2(f)$ are lower than estimated in Chapter 5. For example, I found in Chapter 5 that one source contributed $\sim 50\%$ of $h_c^2(f)$ at a frequency of $2f_{\rm yr}/3$ in a frequency bin of width $(5\,{\rm yr})^{-1}$ (see Figure 5.2). In contrast, the empirical modelling in this Chapter suggests that the strongest ~ 400 sources in such a frequency bin contribute $\sim 50\%$ of $h_c^2(2f_{\rm yr}/3)$.

The models considered in this Chapter and in Chapter 5 clearly predict different numbers of the most massive binary SMBHs. While this discrepancy will only be resolved with GW observations, I point out that the Schechter functions for the GSMFs that I use under-predict observed galaxy counts at the highest masses and the low-

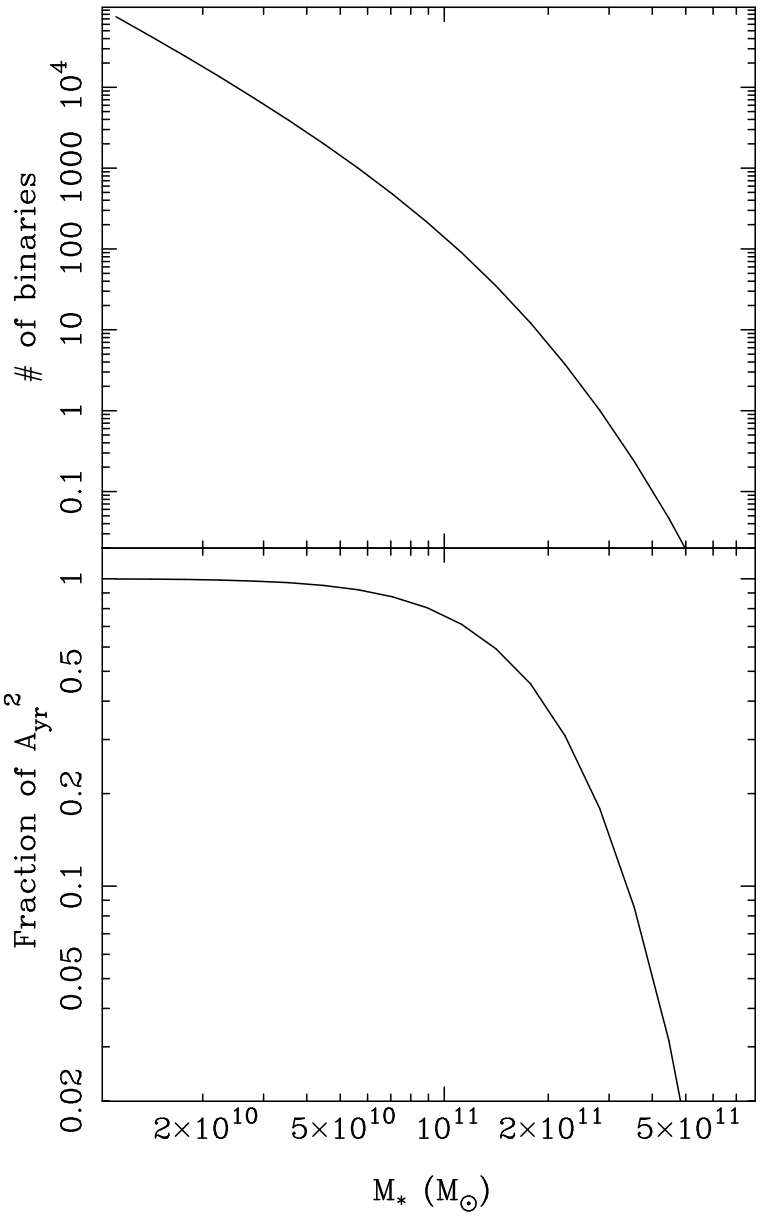

Figure 7.5: *Top:* The numbers of binary SMBH sources predicted by the fiducial model radiating at a GW frequency of $f_{\rm yr}$ in a frequency bin of width $\Delta f = (10\,{\rm yr})^{-1}$ at and above given values of M_*. *Bottom:* The fractions of $A^2_{\rm yr}$ contributed by binary SMBHs at and above given values of M_*.

est redshifts (Baldry et al. 2012; Muzzin et al. 2013). Hence, it is possible that my model under-represents the contributions of the most massive binary SMBHs to the total GW signal. Differing typical galaxy merger mass ratios in cluster and field environments (e.g., Lotz et al. 2013) are a further complicating factor.

7.6 Conclusions

In this Chapter, I predicted the strength of the GWB from binary SMBHs and the occurrence of individual binary SMBH GW sources. My approach was to use a selection of recent observational estimates for the average times between major mergers for galaxies with $M_* > 10^{10} M_\odot$ and $z < 3$ and for the GSMFs of early- and late-type galaxies in this mass and redshift range. I combined these quantities with empirical relations between galaxy and bulge stellar masses and between bulge and SMBH masses.

I find that while current PTAs are unlikely to be sensitive to individual binary SMBHs, a PTA consisting of ∼100 pulsars timed with ∼100 ns precision for 5 yr will be sensitive to up to ∼ 3 binary SMBHs. Such a PTA may be achievable with the SKA (Lazio 2013), but is possibly beyond the capabilities of FAST (Hobbs et al. 2014). Even such a PTA will, however, have a less than 0.1% chance of detecting a GW memory burst from a coalescing binary SMBH. Thus, I conclude that while individual binary SMBHs may be detectable with a PTA based on next-generation radio telescopes, memory bursts from coalescing SMBHs are not likely to be detectable with any envisaged PTA. I caution, however, that my model may under-represent the numbers of individual bright GW sources relative to cosmological-scale simulations of the galaxy and SMBH populations.

I predict that the characteristic strain amplitude of the GWB lies in the range $5.1 \times 10^{-16} < A_{\rm yr} < 2.4 \times 10^{-15}$ with 95% confidence, accounting for a variety of uncertainties. The upper end of the predicted amplitude range is equivalent to the best published 95% confidence upper limit on the GWB amplitude (Shannon et al. 2013).

The dominant uncertainty in predicting the GWB amplitude appears to be caused by differences in theoretical predictions for the major merger timescale of massive galaxies. Higher values within the predicted range for $A_{\rm yr}$ correspond to the more commonly preferred choice of galaxy merger timescale (L08); GWB amplitudes $A_{\rm yr} > 10^{-15}$ are within the sensitivity ranges of current and future PTAs. I

strongly urge further work on quantifying the galaxy merger timescale, in particular for the mergers between massive early-type galaxies at redshifts $z < 1.5$ which are likely to host the dominant contributors to the GWB. The other significant uncertainty in the predictions is in the local form and possible redshift-evolution of the $M_\bullet - M_{\rm bul}$ relation. PTA upper limits on or detections of the GWB may be able to meaningfully improve knowledge of such otherwise poorly constrained facets of the formation and evolution of galaxies and SMBHs.

Chapter 8

PPTA tests of models for the GWB from binary SMBHs

I compare a range of scenarios for the gravitational-wave background from binary supermassive black holes (SMBHs) with the pulsar timing array (PTA) results presented in Chapter 4. I consider predictions from this thesis as well as from other studies. I find that the scenario for the binary SMBH population put forward by McWilliams et al. (2014) is possibly inconsistent with current PTA measurements. Based on the modelling of Chapter 7, I derive a constraint on the merger timescale of massive galaxies: $\tau_m > 0.1$ Gyr with 95% confidence. This Chapter describes my original work, which has partly been published (Shannon et al. 2013).

8.1 Introduction

It is clear from the results of Chapter 7 that constraints on a GWB currently represent the best means of testing models for the binary SMBH population with PTAs. PTA constraints on GWs from individual binary SMBHs, either those with wide orbits or those in the moments of coalescence, do not appear to be close to being inconsistent with current models for source counts (see also Sesana et al. 2009). However, the PPTA 95% upper limit on the amplitude of the characteristic strain spectrum of the GWB (parameterised as $h_c(f) = A_{\rm yr}(f/f_{\rm yr})^{-2/3}$), $A_{\rm yr} < 2.4 \times 10^{-15}$, already appears to exclude some scenarios for the binary SMBH population. For example, McWilliams et al. (2014) predict $A_{\rm yr} = 3.1 \times 10^{-15}$ with 0.2 dex uncertainty

(1σ) under the assumption that all evolution in the galaxy stellar mass function at redshifts $z < 1$ is driven by galaxy mergers.

I have previously considered the question of what astrophysical results may follow from PTA constraints on or detections of a GWB identified with binary SMBHs. The results in Chapter 6 suggest that the GWB characteristic strain spectrum takes the standard power-law form (Equation 5.1) for frequencies $f \gtrsim 10^{-8}\,\mathrm{Hz}$, with the degree of attenuation relative to this form at lower frequencies being dependent on the strength of the coupling between binary SMBHs and their environments. Binary SMBHs are hence likely to be evolving solely under GW emission when radiating at higher frequencies. If binary SMBHs do not stall at wide orbits, this implies that the amplitude of the GWB spectrum at higher frequencies, parameterised as A_{yr}, is directly related to the coalescence rate of binary SMBHs.

The GWB is likely to be dominated by major mergers between massive early-type galaxies at redshifts $z \lesssim 1.5$. These mergers are rare and difficult to identify observationally, with typical massive early-type galaxy pair fractions of $\lesssim 7\%$ for $z < 1$ (e.g., López-Sanjuan et al. 2012), and all techniques for the identification of bound binary SMBHs rely on the presence of significant amounts of gas surrounding the SMBHs which are unlikely to exist in such systems. There is also significant debate as to exactly how much of the evolution in the massive early-type galaxy population at $z \lesssim 2$ is driven by galaxy mergers (Conselice 2014, and references therein). PTA studies of the GWB could contribute substantially to this field. In general, as was discussed in Chapter 7, tests of a simple specification for the binary SMBH population may constrain basic properties of the evolution of the galaxy and SMBH population from a largely empirical standpoint. Further, PTA results are a further constraint for specific models for the GWB amplitude, such as the Guo et al. (2011) semi-analytic galaxy formation model implemented in the Millennium dark matter simulation (Springel et al. 2005).

In this Chapter, I calculate the probabilities of different models for the binary SMBH population being valid, including those presented in this thesis and some others (Sesana 2013b; Kulier et al. 2013; McWilliams et al. 2014). These models are briefly described in §8.2. I compare the model predictions for the GWB characteristic strain spectrum with the PPTA constraints described previously, utilising constraints derived assuming both Gaussian (see §4.1) and non-Gaussian (see §5.9) GWB-induced ToA variations. I also discuss the implications of the constraints

being placed at a GW frequency $f_{\rm PPTA} = 2.8 \times 10^{-9}\,{\rm Hz}$. Based on the work in Chapter 7, I present a preliminary constraint on characteristic timescale, τ_m, for major massive galaxy mergers.

8.2 Summary of GWB models

In this thesis, I have discussed two separate models for the coalescence rate of binary SMBHs, including one based on the Guo et al. (2011, hereafter G11) semi-analytic galaxy formation model and the Millennium simulation (Springel et al. 2005), and one based on empirical estimates of the galaxy merger rate and the relations between SMBH masses and galaxy masses for different galaxy types. The prediction for the GWB amplitude resulting from the G11 model, presented in §5.8, has a narrow uncertainty range, which represents uncertainty in the SMBH-mass scale factor given recent measurements of SMBH and bulge masses. A range of values of the other free parameters of the G11 model are possible given other uncertainties in other observational constraints (Mutch et al. 2013), likely resulting in a larger possible range of predicted GWB amplitudes. I demonstrated the effects of such uncertainties, including the effects of binary orbital decay driven by stellar environments and possible binary eccentricities in Chapter 6. Here, I compare both the fiducial prediction and the broader prediction for the GWB that I derive from the G11 model.

The model presented in Chapter 7 results in a wide range of predicted GWB amplitudes, which is intended to encompass numerous uncertainties in specifying the SMBH-SMBH coalescence rate. However, as previously discussed, it is not clear that the possible ranges of some of the input parameters, in particular the timescale over which major galaxy mergers are observable, are fully understood. Hence, I discuss a constraint on this timescale given prior assumptions for the uncertainty ranges of all other input parameters.

In Chapter 7, I also introduced three independent predictions for the GWB amplitude which are consistent, in particular, with recent SMBH mass measurements (Sesana 2013b; Kulier et al. 2013; McWilliams et al. 2014). All predictions assume circular orbits for all binary SMBHs. I also compare these predictions with the PPTA constraints on the GWB. The models and predictions are as follows.

Sesana (2013b), hereafter in this Chapter S13. This model is quite similar to that presented in Chapter 7 in that the SMBH-SMBH coalescence rate is specified using empirical results. However, S13 uses a larger selection of results than I consider, including less accurate estimates of the galaxy stellar mass function and the fraction of galaxies undergoing mergers. Also, S13 considers every type of galaxy merger (i.e., early-type with late-type, early-type with early-type, etc.) to be equally possible. Finally, rather than generating Monte Carlo realisations of possible GWB amplitudes, S13 simply considers maximum-likelihood values and lower and upper limits for each input quantity to be equally possible, and evaluates $A_{\rm yr}$ for all combinations of inputs. The resulting 1σ range of the GWB amplitude is $5.6 \times 10^{-16} < A_{\rm yr} < 2.1 \times 10^{-15}$; the posterior probability distribution for $A_{\rm yr}$ is lognormal.

Kulier et al. (2013), hereafter in this Chapter K13. Using results from previous high-resolution cosmological hydrodynamical simulations of a cluster and void region selected from a larger-scale, low-resolution run (Cen 2011), K13 presented the first prediction of the GWB amplitude based only on numerical simulations of baryons and dark matter. A sophisticated scheme to grow SMBHs through both hot and cold gas accretion, as well as mergers, was employed, and associated feedback processes, post-coalescence SMBH recoil and binary SMBH stalling were accounted for. The GWB amplitude was calculated by averaging over the binary SMBH populations in the cluster and void regions, which may cause some bias may be present in this estimate. The resulting GWB amplitude was $A_{\rm yr} = 1.5 \times 10^{-15}$, with a 0.13 dex 1σ uncertainty.

McWilliams et al. (2014), hereafter in this Chapter M14. This novel model for the binary SMBH population was built on the assumption that all evolution in the galaxy stellar mass function for redshifts $z < 1$ is driven by galaxy mergers, with negligible star formation. This approach reproduces the growth in the number of galaxies at masses $M_* \gtrsim 10^{10} M_\odot$, and represents a maximal galaxy merger rate. M14 also assume that all galaxies are bulge dominated. The predominant contribution of bulge-dominated galaxies to the GWB, demonstrated in Chapter 7, means that this assumption is unlikely to significantly raise the GWB amplitude prediction. Ultimately, M14 predict $A_{\rm yr} = 3.1 \times 10^{-15}$ with 0.2 dex 1σ uncertainty, with a possible low-frequency

turn-over consistent with the results in Chapter 6 assuming a binary SMBH orbital hardening rate in a fixed stellar background consistent with the results of Quinlan (1996). I have scaled the prediction by a factor of $(3/4)^{1/2}$ to correct the expression used by M14 for the orientation-averaged rms GW strain amplitude for binary SMBHs.

8.3 Testing GWB models with the PPTA

8.3.1 The probabilities of GWB models given PPTA constraints

I compare model predictions with the PPTA constraints (assuming Gaussian and non-Gaussian GWB-induced pulse arrival time variations) at the GW frequency $f_{\rm PPTA}$ in terms of $\Omega_{\rm GW}$ evaluated at this frequency. The GWB characteristic strain spectrum may be related to $\Omega_{\rm GW}(f)$ through Equations 3.42 and 3.46. Let the predictions from a model, denoted M, be represented by the posterior predictive probability density function $\rho_M(\Omega_{\rm GW})$. This can be interpreted as the conditional probability density function of the respective model being true given a value of $\Omega_{\rm GW}$:

$$\rho_M(\Omega_{\rm GW}) = \frac{d\Pr(M|\Omega_{\rm GW})}{d\log\Omega_{\rm GW}}. \tag{8.1}$$

Now, in the process of evaluating 95% confidence upper limits on the GWB amplitude with the PPTA data (see §4.1), simulated distributions of the detection statistic $\hat{A}$ given values of $A_{\rm yr}$ were compared with the measured value of the detection statistic from the data. The 95% confidence upper limits were taken as the value of $A_{\rm yr}$ such that 95% of the simulated values lay above the measured value and were hence inconsistent with it. Here, for a given $A_{\rm yr}$, I interpret the fraction of simulated $\hat{A}$ values lying above the measured value as the probability of $A_{\rm yr}$ being consistent with the PPTA data, $\Pr(\Omega_{\rm GW})$.

Given $\rho_M(\Omega_{\rm GW})$ and $\Pr(\Omega_{\rm GW})$, the law of total probability implies

$$\Pr(M) = \int_{-\infty}^{\infty} \rho_M(\Omega_{\rm GW})\Pr(\Omega_{\rm GW}) d\log\Omega_{\rm GW}. \tag{8.2}$$

The forms of $\Pr(\Omega_{\rm GW})$ assuming both Gaussian and non-Gaussian GWB statistics

Table 8.1: Probabilities of different GWB models given PPTA data.

Model description	$\Pr(M)$
G11, fiducial	0.38
G11, non-Gaussian	0.51
G11, environments	0.9
S13	0.54
K13	0.39
M14	0.07
M14, non-Gaussian	0.09

are well modelled by a complementary Gaussian error function in logarithmic units. For a Gaussian GWB,

$$\Pr(\Omega_{\rm GW}) = \int_{\log\Omega_{\rm GW}}^{\infty} \frac{1}{\sqrt{2\pi(0.25)^2}} \exp\left(\frac{-(\log\Omega'_{\rm GW}+9.37)^2}{2(0.25)^2}\right) d\log\Omega'_{\rm GW}, \quad (8.3)$$

and for a non-Gaussian GWB,

$$\Pr(\Omega_{\rm GW}) = \int_{\log\Omega_{\rm GW}}^{\infty} \frac{1}{\sqrt{2\pi(0.33)^2}} \exp\left(\frac{-(\log\Omega'_{\rm GW}+9.34)^2}{2(0.33)^2}\right) d\log\Omega'_{\rm GW}. \quad (8.4)$$

In this subsection, I evaluate $\Pr(M)$ for the models discussed above given $\Pr(\Omega_{\rm GW})$ derived assuming Gaussian GWB statistics and $\rho_M(\Omega_{\rm GW})$ derived assuming the standard power-law form for the GWB characteristic strain spectrum. Values of $\Pr(M)$ for the fiducial G11 prediction and the predictions of S13, K13 and M14 are listed in Table 8.1. In the Table, I also demonstrate the effects of assuming non-Gaussian GWB statistics (labelled 'G11, non-Gaussian' and 'M14, non-Gaussian') and of binary SMBH environmental-driven attenuation of the GWB (labelled 'G11, environments') in further calculations of $\Pr(M)$ for the G11-based model and the M14 model. In Figure 8.1, I show $\Pr(\Omega_{\rm GW})$ from Equations 8.3 and 8.4, as well as $\rho_M(\Omega_{\rm GW})$ for each of the models I consider.

The PPTA data assuming a Gaussian GWB and no attenuation at low frequencies because of binary SMBH environments are inconsistent with the M14 prediction at the 93% confidence level. The effect of assuming a non-Gaussian GWB is to increase the variance of the simulated distributions of the detection statistic $\hat{A}$ for a given $A_{\rm yr}$, which, as I have shown, leads to the 95% confidence upper limit on $A_{\rm yr}$

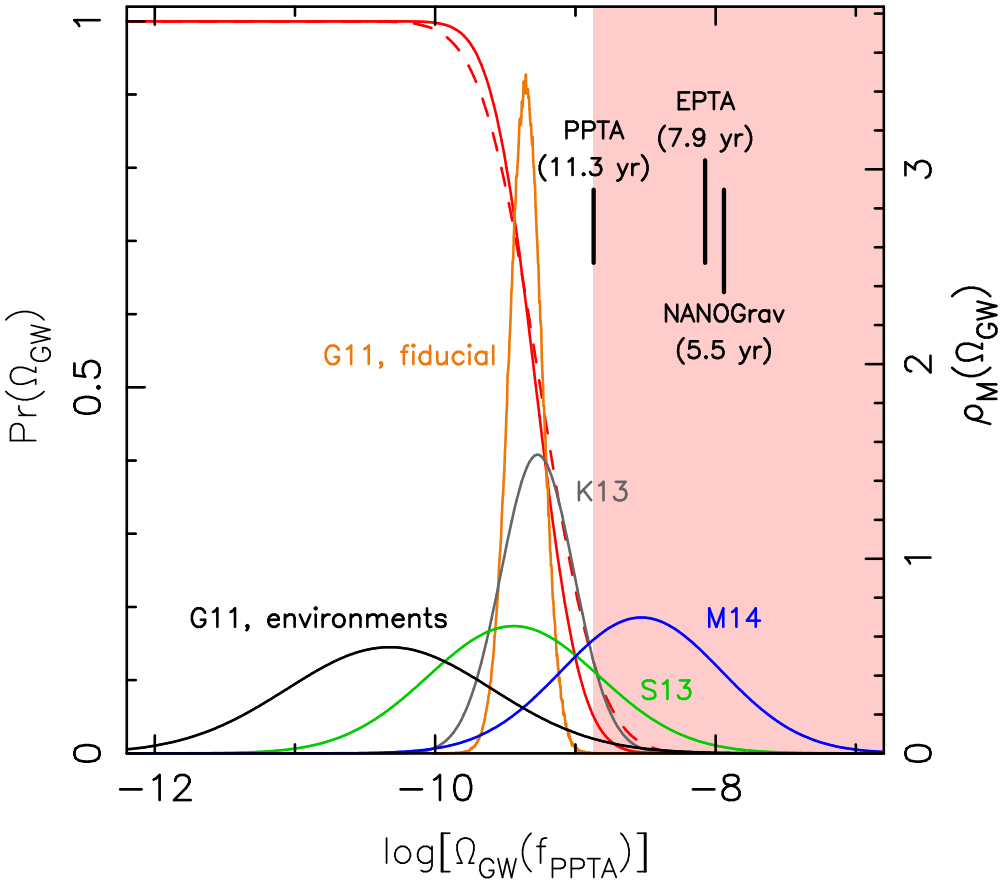

Figure 8.1: The left-hand ordinate represents the PPTA probabilities of different values of $\Omega_{\rm GW}(f_{\rm PPTA})$, $\Pr(\Omega_{\rm GW})$, which I display assuming both Gaussian (red solid curve) and non-Gaussian (red dashed curve) statistics for the GWB. The right-hand ordinate represents the posterior predictive probability density functions, $\rho_M(\Omega_{\rm GW})$, for a selection of models, as listed in Table 8.1. The function $\rho_M(\Omega_{\rm GW})$ used in calculating the probability, $\Pr(M)$, for the 'G11, non-Gaussian' model in Table 8.1 is the same as that labelled 'G11, fiducial'. I also show the values of $\Omega_{\rm GW}$ that are inconsistent with the PPTA data with greater than 95% confidence as a salmon shaded region assuming a Gaussian GWB, along with the 95% confidence upper limits on $\Omega_{\rm GW}$ from the PPTA, EPTA (van Haasteren et al. 2011) and NANOGrav (Demorest et al. 2013) PTA groups (the latter two assuming a standard power-law GWB characteristic strain spectrum) as well as the GW periods of maximum sensitivity for each upper limit.

increasing from 2.4×10^{-15} to 2.7×10^{-15}. In the case of a non-Gaussian GWB, the fiducial prediction based on the G11 model is 13% less inconsistent with the PPTA data, while the M14 prediction is 3% less inconsistent. Further degradation in this result for the G11 model, by 39%, occurs when the model for environment-driven attenuation, as well as larger predictive uncertainty, are included.

8.3.2 A lower limit on the massive galaxy major merger timescale

Numerical galaxy merger simulations by Lotz et al. (2008) and Lotz et al. (2010), hereafter collectively L08, found that the average timescale for interacting galaxy pairs to linger between projected separations of $5\,h^{-1}\,\mathrm{kpc}$ and $20\,h^{-1}\,\mathrm{kpc}$[1] was $\tau_m = 0.3\,\mathrm{Gyr}$. This was approximately three times smaller than the timescale found by Kitzbichler & White (2008) by averaging over mergers from a semi-analytic model implemented in the Millennium simulation, and assuming the standard dynamical friction formula and circular galaxy orbits. Various caveats to both results exist, as discussed in Chapter 7, which could mean that the uncertainty in predicting τ_m is even greater than the range spanned by the predictions of L08 and Kitzbichler & White (2008). This uncertainty caused the most variation in my empirical prediction of A_{yr} in Chapter 7 given all current uncertainties.

By simulating 1000 realisations of A_{yr} given all uncertainties discussed in Chapter 7, but assuming $\tau_m = 0.3\,\mathrm{Gyr}$ (L08), I found that $A_{\mathrm{yr}} = 1.5 \times 10^{-15}$ with 0.12 dex 1σ lognormal uncertainty. Furthermore, I confirmed that $A_{\mathrm{yr}} \propto \tau_m^{-1/2}$, allowing a definition of the probability density function in $\Omega_{\mathrm{GW}}(f_{\mathrm{PPTA}})$, $\rho_{\tau_m}(\Omega_{\mathrm{GW}})$, for any τ_m. Then, Equation 8.2 can be used to evaluate the probability $\Pr(\tau_m)$. I show the results for various scale factors $\tau_m/(0.3\,\mathrm{Gyr})$ in Figure 8.2. The resulting 95% confidence constraint is $\tau_m/(0.3\,\mathrm{Gyr}) > 1/3$.

8.4 Discussion and conclusions

Under the optimistic, albeit standard, assumptions of Gaussian GWB-induced pulsar ToA variations and negligible environmental attenuation of the GWB at

[1]The choice of projected separation range is not particularly important, as long as the same range is used to compare different predictions.

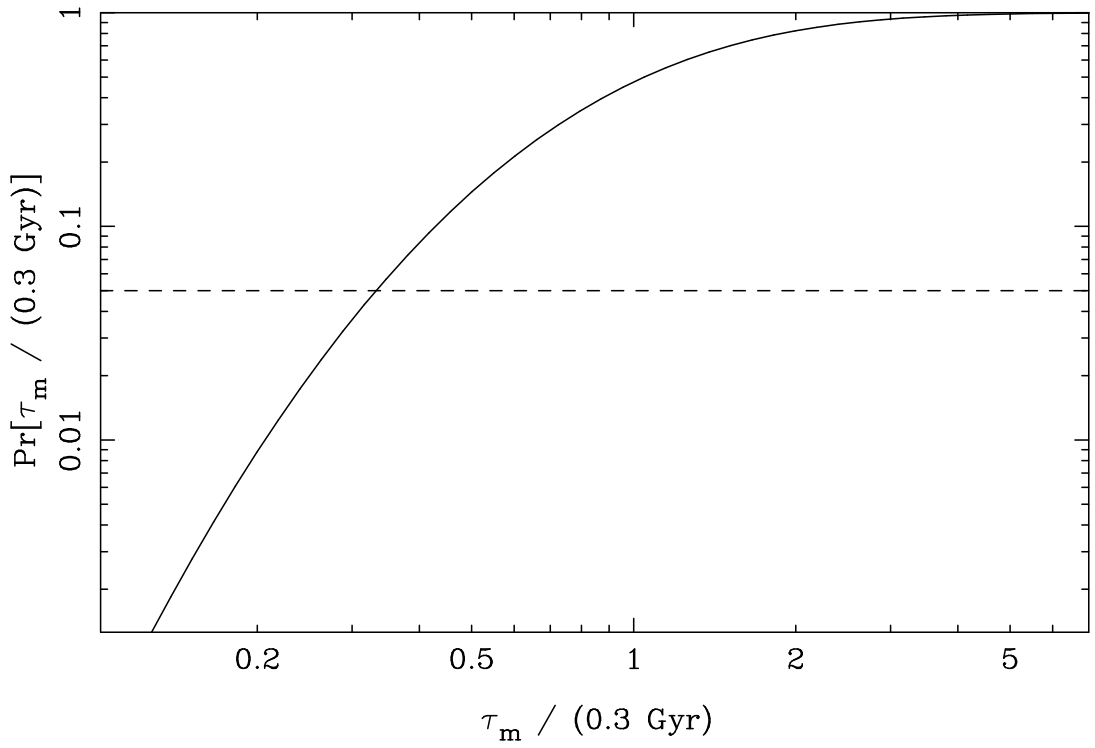

Figure 8.2: The probabilities $\Pr[\tau_m/(0.3\,\mathrm{Gyr})]$ of different values of $\tau_m/(0.3\,\mathrm{Gyr})$ given the PPTA results, shown as a solid curve. The dashed horizontal line indicates $\Pr[\tau_m/(0.3\,\mathrm{Gyr})] = 0.05$.

$f_{\mathrm{PPTA}} = 2.8 \times 10^{-9}\,\mathrm{Hz}$, a selection of scenarios for the coalescence rate of binary SMBHs are inconsistent with the current PPTA data set. The M14 model is inconsistent at the 93% confidence level. In general, I interpret the probabilities $\Pr(M)$ for each model as the fractions of possible scenarios predicted by each model which are consistent with the PPTA data. For example, given 100 universes with a binary SMBH coalescence rate described by a random combination of input parameters to the S13 model, and further given exactly the same PPTA data sets, I expect 54 of these universes to result in true values of the detection statistic $\hat{A}$ which are consistent with the measured value. In contrast, the naïve interpretation of the probabilities $\Pr(M)$ as representing the fractions of possible permutations of the models is incorrect. Nonetheless, it is striking that some *otherwise plausible* scenarios for the coalescence rate of binary SMBHs appear to be inconsistent with the PPTA data set.

I have further shown that the major merger timescale of massive galaxies, normalised to galaxy separations between $5\,h^{-1}\,\mathrm{kpc}$ and $20\,h^{-1}\,\mathrm{kpc}$, is greater than 100 Myr with 95% confidence. Given that the GWB is likely dominated by merg-

ers between massive ($M_* \gtrsim 10^{10.5} M_\odot$) early-type galaxies at redshifts $z \lesssim 1$, the constraint is best discussed in that context. Various authors infer that significant fractions of the evolution in the massive early-type galaxy population over these redshifts is merger-driven, with rare major mergers and numerous minor mergers playing similarly important roles (e.g., Xu et al. 2012; López-Sanjuan et al. 2012). Constraining the galaxy merger timescale to be less than three times faster than that suggested by L08 constrains the merger rate to be less than three times as great as what is currently measured. The work of López-Sanjuan et al. (2012) suggests that a merger timescale that is even 1.5 times faster than that of L08 results in the entirety of the observed size evolution of early-type galaxies being accounted for by major mergers. In general, while the constraint I place on τ_m is interesting, it is likely that better PTA data are required before more meaningful results can be obtained.

I have shown in this thesis that both the aforementioned standard assumptions for the GWB may not be correct. However, the effects of assuming a non-Gaussian GWB appear to be minor in determining the probabilities $\Pr(M)$; the probability of inconsistency between the PPTA data and the M14 model reduces by 2% given in the non-Gaussian case. Low-frequency attenuation of the GWB because of binary SMBH losses of energy and angular momentum to environments appears to be a more significant issue. Given the modelling in Chapter 6, the probability of inconsistency between the PPTA data and the G11-based prediction reduces by 52%. While this result is somewhat exaggerated by the larger number of uncertainties included in the prediction made in Chapter 6, the problem remains; using near-identical assumptions for binary SMBH stellar environments to those in Chapter 6, M14 also suggest that the GWB is attenuated by $f \lesssim 10^{-8}$ Hz. As discussed in Chapter 6, it appears that this particular model for binary environments (Sesana 2010) may result in faster orbital decay than more realistic numerical models which self-consistently trace dry galaxy mergers (Khan et al. 2012), possibly inflating the maximum GWB attenuation frequency by ~ 0.4 dex. Further work on tracing the formation and evolution of binary SMBHs in major mergers of early-type galaxies is required to better understand the low-frequency attenuation of the GWB.

Chapter 9

Conclusions

My chief motivation in pursuing the research described in this thesis was to investigate the cosmological populations of galaxies and supermassive black holes (SMBHs) using pulsar timing arrays (PTAs). When I began this research, it was already known that binary SMBHs in galaxy merger remnants would emit gravitational waves (GWs) that could be detectable with sufficiently sensitive PTAs. However, the scientific potential and feasibility of this endeavour were not well understood. At the same time, the Parkes PTA (PPTA) collaboration was assembling a data set with significantly improved sensitivity to GWs from binary SMBHs (Manchester et al. 2013). This data set raised the compelling possibility, for the first time, of deriving meaningful constraints on the populations of galaxies and SMBHs using GWs.

In this thesis, I therefore sought to carefully characterise expectations for the collated GW signals from binary SMBHs. Having assembled this theoretical base, I then considered the implications of the best existing PTA-based upper limits on astrophysical GW signals. In this Chapter, I first précis my expectations for the nature and detectability of GWs from binary SMBHs, and conclude with my view on current and possible future PTA-driven insights into the assembly histories of galaxies and SMBHs.

9.1 The *mare incognitum* of gravitational waves awaiting PTAs

At some point during the merger of a pair of massive galaxies, the central SMBHs (if present) are likely to form a gravitationally-bound binary system. Such a binary system will lose energy and angular momentum first to its stellar or gaseous environment, and then through the emission of GWs, which eventually leads to the SMBHs coalescing. I have shown (as detailed in Chapter 7) that, if binary SMBHs form and coalesce ubiquitously in major mergers of massive galaxies, the total energy density in GWs from the population of binary SMBHs may be sufficiently large so as to be detectable with PTAs, likely over the next decade.

In making this prediction, I used a suite of observational data to directly specify the SMBH-SMBH coalescence rate, and also assumed that all binaries evolve under GW emission alone. I showed that the characteristic strain amplitude of the GW background (GWB) generated by all binary SMBHs at a frequency of $(1\,\mathrm{yr})^{-1}$ lies in the range $5.1\times10^{-16} < A_{\mathrm{yr}} < 2.4\times10^{-15}$ with 95% confidence. The preferred prescription for the timescale over which galaxy mergers occur led to a systematically higher prediction: $9\times10^{-16} < A_{\mathrm{yr}} < 2.7\times10^{-15}$ with 95% confidence. While it is not straightforward to predict the future sensitivities of PTAs to a GWB from binary SMBHs, Siemens et al. (2013) suggested that a GWB of amplitude $A_{\mathrm{yr}} = 10^{-15}$ will indeed be detectable by the year 2021 using observations from the North American Nanohertz Observatory for GWs (NANOGrav).

Besides the GW background (GWB) generated by the cosmological population of binary SMBHs, the detection of GWs from individual binary SMBHs has engendered immense interest among the PTA community. In particular, the electromagnetic identification of a GW source would provide the ultimate evidence for the reliability of PTA searches for GWs, and allow for fascinating investigations into General Relativity and accretion physics. However, even my most optimistic predictions for the numbers of individual binary SMBHs emitting continuous-wave GWs of different strengths suggest that a PTA data set consisting of 100 pulsars timed with 100 ns precision over 5 yr would be required to expect a detection. Obtaining such a data set is unfortunately at the edge of the foreseen capabilities of currently planned radio telescopes, such as the Five hundred metre Aperture Spherical Tele-

scope (FAST; Li et al. 2013; Hobbs et al. 2014) and the Square Kilometre Array (SKA; Cordes et al. 2004; Lazio 2013). My modelling also suggests that bursts of GWs from the closest approaches of individual eccentric binary SMBHs, and GW bursts with 'memory' from SMBH-SMBH coalescence events, are unlikely to ever be viable sources for PTAs.

The uncertainty in my predictions for the GWB amplitude are dominated by quantities for which little or no observational information currently exists. These quantities include the galaxy merger timescale, and to a lesser extent the SMBH-galaxy scaling relations both locally and at higher redshifts. The galaxy merger timescale has only been derived through numerical (e.g., Lotz et al. 2008) or semi-analytic (e.g., Kitzbichler & White 2008) modelling, and existing predictions are quite varied. All observational constraints on the redshift-evolution of the SMBH-galaxy scaling relations are based on observations of quasars, and are hence subject to a number of possible biases (e.g., Kormendy & Ho 2013). However, of all binary SMBHs, those formed in the mergers of particularly massive ($M_* \gtrsim 10^{10.5}$) early-type galaxies at reasonably low redshifts ($z \lesssim 1.5$) appear to be the dominant contributors to the GWB. Efforts towards identifying galaxy merger timescales specifically for such systems would hence be of particular use in refining predictions of the GWB amplitude.

While the predictions summarised above were based on an empirical characterisation of the binary SMBH population, a more comprehensive inquiry into the properties of this population required a different approach. I used a semi-analytic model for galaxy and SMBH formation (Guo et al. 2011) implemented in the Millennium dark matter numerical simulation (Springel et al. 2005) to predict the SMBH-SMBH coalescence rate. This method has the advantage of providing a simulated sample of binary SMBHs at various redshifts large enough to be free of biases related to observations and cosmic variance. The galaxies hosting binary SMBHs were also specified. However, the key disadvantage of this approach is that the resulting predictions are specific to given values of the free parameters of the model, as well as to a large set of model assumptions. While the model is, of course, tuned to observations, understanding the possible ranges of the free parameters given observational uncertainties is non-trivial. Further, while the model assumptions are physically motivated, they are not all strongly evidenced in themselves. For example, a specific galaxy merger timescale based on the Chandrasekhar (1943) dynamical friction

formula is assumed by Guo et al. (2011), which, for major mergers, is not consistent with detailed numerical simulations (e.g., Lotz et al. 2008). These issues are, however, not significantly relevant to my main motivations for adopting this physically-motivated modelling approach, the conclusions from which are summarised below (and elaborated upon in Chapters 5 and 6).

The large simulated sample of binary SMBHs derived from the Guo et al. (2011) model implemented in the Millennium simulation was uniquely useful in analysing the statistics of pulsar time of arrival (ToA) variations induced by GWs from the population of binary SMBHs. I showed that these statistics are mildly non-Gaussian, contrary to what was previously assumed. Specifically, at every GW frequency accessible with a 5 yr PTA data set, the measured power-spectral densities of realisations of ToA-variation time series displayed greater variance than if the ToA variations were assumed to correspond to a Gaussian random process. Assuming these statistics for the ToA variations induced by the GWB led to a $\sim 10\%$ degradation in the 95% confidence upper limit on the GWB amplitude derived from Parkes PTA data (Manchester et al. 2013) by Shannon et al. (2013).

The predicted non-Gaussian GWB-induced ToA variations were caused by the GWB being dominated by a few bright individual GW sources, under this specific model for the binary SMBH population. The contributions of the brightest few sources to the GWB were, however, found to be smaller in my empirically-derived model. While this discrepancy will likely be only resolved with GW observations, it may be possible that the most massive binary SMBH systems are slightly under-represented in the empirically-derived model.

The efficiency of the loss of energy and angular momentum from binary SMBHs to their environments is another unknown quantity which affects, in particular, predictions of the spectrum of the GWB at low frequencies. By coupling a specific model for binary orbital evolution (Sesana 2010) in stellar environments with the physically-motivated approach towards predicting the binary SMBH population, I showed that fewer-than-expected binary SMBHs radiating GWs within the PTA frequency band may be present. I predicted a resulting attenuation in the GWB characteristic strain spectrum with respect to the standard power-law form ($h_c(f) \propto f^{-2/3}$) at frequencies of $f \lesssim 10^{-8}\,\mathrm{Hz}$. Allowing for the possibility of binary orbital eccentricities, which are likely to be grown through environmental interactions, led to further low-frequency attenuation of the spectrum because of the redistribution of

emitted GW power to higher frequencies. While the binary orbital evolution model that I assumed may overestimate the efficiency of binary-environment coupling, such that the turn-over frequency may instead be $\sim 3.5 \times 10^{-9}$ Hz, PTAs are likely to be less sensitive to the GWB at low frequencies than currently thought. Nonetheless, while the caveats of non-Gaussian GWB-induced ToA variations and coupling between binary SMBHs and their environments need to be accounted for, detecting and performing science with the stochastic GWB from binary SMBHs remains the most promising prospect for PTAs.

9.2 Advancing galaxy and SMBH astrophysics

Over the last 50 years, extraordinary efforts have been directed at understanding the discovery scopes of GW detection experiments. This work has been essential in shaping the designs of these experiments; for example, the Advanced LIGO facility (Harry & LIGO Scientific Collaboration 2010), which will commence observations in mid-2015, is the first such experiment that is claimed to be *likely* to detect GWs during its period of operation.

My intent for this thesis, however, was to transcend such a technical guide for PTAs, and attempt to draw scientific conclusions about the population of binary SMBHs from the most recent PTA data. This Chapter has so far collated my inputs towards characterising the GW signals from binary SMBHs that lie within the reach of PTAs. While searches for GWs from individual binary SMBHs, in particular from candidate binary systems (e.g., Jenet et al. 2004), are of clear importance, the comprehensive approach adopted herein has favoured investigating the collective GWB generated by all binary SMBHs. I have shown that PTA data sets are likely to be more sensitive to the GWB than to individual sources. Therefore, constraining and eventually detecting a GWB is likely to be the best avenue for population-wide statements on the assembly histories of galaxies and SMBHs.

In Chapter 8, I calculated the probabilities of different models for the binary SMBH population being inconsistent with the best existing PTA constraints (Chapters 4 and 5; Shannon et al. 2013). I assumed that all binary SMBHs are in circular orbits and evolving only under the emission of GWs, specifically when radiating at the frequency of $f_{\rm PPTA} = 2.8 \times 10^{-9}$ Hz at which the constraint was set. Even assuming non-Gaussian statistics for the GWB-induced ToA variations, I found that the

McWilliams et al. (2014) model is inconsistent with the data at the 91% confidence level. However, it is important to recognise that the GWB could be significantly attenuated at $f_{\rm PPTA}$ if coupling between binary SMBHs and their environments is strong, and further attenuated if large orbital eccentricities are grown through environmental interactions.[1]

None of the other models for the binary SMBH population that I considered in Chapter 8, including those presented in this thesis, resulted in as high a GWB amplitude as was predicted by McWilliams et al. (2014). This was because McWilliams et al. (2014) assumed that (*a*) all evolution in the stellar mass function of massive galaxies for redshifts $z < 1$ is caused by galaxy mergers, and (*b*) that all massive galaxies consist only of bulges. In contrast, I assumed that the merger rate of massive galaxies is described in an unbiased manner by observations, and that massive galaxies consist of a combination of late-type systems, lenticulars and ellipticals. In general, all models for the binary SMBH population that I considered besides McWilliams et al. (2014) resulted in consistent predictions for the GWB amplitude, and none was inconsistent with the PPTA constraint at as high a confidence level.

I also used my empirically-motivated modelling of the binary SMBH population to demonstrate how a constraint on the galaxy merger timescale may be obtained using PTA data. I found $\tau_{\rm m} > 0.1\,{\rm Gyr}$ with 95% confidence for the major mergers of massive galaxies, where $\tau_{\rm m}$ is normalised to the time spent by merging systems between projected separations of $20\,h^{-1}\,{\rm kpc}$ and $5\,h^{-1}\,{\rm kpc}$. This result is reliant on the standard assumptions for the GWB from binary SMBHs, and on a complete specification of the full ranges of other inputs to characterising the binary SMBH population. While the constraint is not in opposition to current models which predict $\tau_{\rm m}$, my work demonstrates how more useful results may be derived in the future as more sensitive PTA observations become available.

To take a broader view of these findings, it is remarkable in itself that we are at a stage where any meaningful constraints on the assembly of the galaxy and SMBH populations can be achieved through searches for GWs. To me, this is hugely encouraging and represents an important achievement towards establishing the credence of performing GW astrophysics with PTAs.

[1] This was also pointed out by McWilliams et al. (2014).

9.3 Studying the GWB with PTAs of the future

The path towards improved PTA data sets is clear. The current PTA collaborations (Kramer & Champion 2013; Manchester et al. 2013; McLaughlin 2013) are expected to conduct observations for the foreseeable future with the largest existing radio telescopes, steadily improving on instrumentation and data analysis methods, and including new millisecond pulsars as they are discovered. The data sets from all PTAs will be combined under the umbrella of the International PTA (Hobbs et al. 2010b) in order to maximise sensitivity to GWs. Currently planned telescopes, such as FAST (Li et al. 2013; Hobbs et al. 2014) in China and the SKA (Cordes et al. 2004; Lazio 2013) mid-frequency component, include PTA observations as a major scientific motivation. Predicting the characteristics of future PTA data sets is, however, not nearly as well-posed a problem as is the case for characterising other GW detection experiments. While pulsars may be reduced in this thesis to utilities for the detection of GWs, they remain astrophysical objects, and ToA measurements are subject to numerous, complex sources of noise (e.g., Cordes & Shannon 2010). For example, the accuracies of ToA measurements are ultimately limited by the stochasticity of individual pulses from pulsars, and not by telescope sensitivity alone.

It is nevertheless likely that over the next few years upper limits on the amplitude of the GWB derived from the IPTA data set will steadily improve. Limits will also be placed at multiple GW frequencies (e.g., Lentati et al. 2013). Eventually, these limits will likely hit a noise floor, which will indicate the presence of either the GWB or pulsar timing noise. If the noise floor is indeed the GWB, confidence in its presence through an analysis of the Hellings & Downs (1983) correlations between different pulsar data sets will slowly increase. New telescopes such as FAST and the SKA will eventually provide substantially better data sets than the IPTA, and may ultimately allow for the GWB amplitude at multiple frequencies to be measured. The question of interest here is: what astrophysics will be possible along this phantasmal observational progression?

In the absence of a detection, upper bounds on the GWB will allow for evermore-interesting tests of models for the binary SMBH population. These tests will be particularly robust and directly relevant to the coalescence rate of SMBH pairs, if (a) the upper bounds are placed at GW frequencies that are demonstrably unaffected

by binary-environment interactions and (b) if non-Gaussian GWB-induced ToA variations are assumed. Based on the work of Kulier et al. (2013), cosmological hydrodynamical simulations such as Illustris (Vogelsberger et al. 2014), which combine baryonic and dark matter components, may predict particularly strong GWBs. These simulations naturally include sophisticated treatments of galaxy merger timescales. More interesting constraints on important predictors of the SMBH-SMBH coalescence rate that have otherwise poor observational information, such as the galaxy merger timescale and the redshift-evolution of the SMBH-galaxy scaling relations, will also be possible with improved GWB limits. Independent progress in narrowing the possible ranges of other such predictors will also aid this endeavour.

A direct measurement of the shape of the GWB spectrum, however fanciful a prospect, would furnish an excellent proving ground for models of binary SMBH interactions with their environments. It is likely that the GWB is dominated by binaries formed in the major mergers of massive early-type galaxies. This restriction lends a useful specificity to modelling binary environments: dynamical friction is likely to be very efficient at forming a new galaxy, which will host very little (cold) gas.

Ultimately, however, the very detection of a GWB would represent strong evidence for the existence of a cosmological population of binary SMBHs emitting GWs, as predicted by the General Theory of Relativity. A measurement of the GWB amplitude at frequencies high enough to be unaffected by binary environmental interactions would further provide an observable which is directly related to the coalescence rate of pairs of SMBHs. This rate is directly dependent on the demographics of SMBHs at various redshifts, the galaxy merger rate, and the efficiency of SMBH-SMBH coalescence in galaxy mergers. Besides enabling investigations into fundamental gravitational physics, PTAs have the potential to uniquely inform our understanding of the growth histories of massive galaxies and SMBHs. Indeed, in this thesis, I have argued that this potential is already being realised.

Bibliography

Abel, T., Bryan, G. L., & Norman, M. L. 2000, ApJ, 540, 39

Acernese, F., Amico, P., Al-Shourbagy, M., et al. 2006, Classical and Quantum Gravity, 23, 63

Agudo, I., Marscher, A. P., Jorstad, S. G., et al. 2012, ApJ, 747, 63

Alcock, C., Allsman, R. A., Alves, D., et al. 1996, ApJ, 471, 774

Alexander, D. M., & Hickox, R. C. 2012, New Astronomy Reviews, 56, 93

Aller, M. C., & Richstone, D. 2002, AJ, 124, 3035

Amaro-Seoane, P., Aoudia, S., Babak, S., et al. 2012, Classical and Quantum Gravity, 29, 124016

Ambartsumian, V. A. 1937, Astronomicheskii Zhurnal, 14, 207

Anholm, M., Ballmer, S., Creighton, J. D. E., Price, L. R., & Siemens, X. 2009, Phys. Rev. D, 79, 084030

Armitage, P. J., & Natarajan, P. 2002, ApJL, 567, L9

Armstrong, J. W. 2006, Living Reviews in Relativity, 9, 1

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., et al. 2014, ArXiv e-prints, arXiv:1404.1267

Babak, S., & Sesana, A. 2012, Phys. Rev. D, 85, 044034

Babcock, H. W. 1939, Lick Observatory Bulletin, 19, 41

Backer, D. C., Kulkarni, S. R., Heiles, C., Davis, M. M., & Goss, W. M. 1982, Nature, 300, 615

Baes, M., Buyle, P., Hau, G. K. T., & Dejonghe, H. 2003, MNRAS, 341, L44

Bahcall, J. N., & Wolf, R. A. 1976, ApJ, 209, 214

Baker, J. G., Centrella, J., Choi, D.-I., Koppitz, M., & van Meter, J. 2006a, PRL, 96, 111102

Baker, J. G., Centrella, J., Choi, D.-I., et al. 2006b, ApJL, 653, L93

Baldry, I. K., Driver, S. P., Loveday, J., et al. 2012, MNRAS, 421, 621

Balick, B., & Brown, R. L. 1974, ApJ, 194, 265

Barkana, R., & Loeb, A. 2001, Physics Reports, 349, 125

Barnes, J. E., & Hernquist, L. 1992, ARA&A, 30, 705

Begelman, M. C. 1978, MNRAS, 184, 53

—. 2002, ApJL, 568, L97

—. 2003, Science, 300, 1898

Begelman, M. C., Blandford, R. D., & Rees, M. J. 1980, Nature, 287, 307

Begelman, M. C., & Rees, M. J. 1978, MNRAS, 185, 847

Begelman, M. C., & Shlosman, I. 2009, ApJL, 702, L5

Begelman, M. C., Volonteri, M., & Rees, M. J. 2006, MNRAS, 370, 289

Benson, A. J. 2010, Physics Reports, 495, 33

Berentzen, I., Preto, M., Berczik, P., Merritt, D., & Spurzem, R. 2009, ApJ, 695, 455

Bernardi, M., Shankar, F., Hyde, J. B., et al. 2010, MNRAS, 404, 2087

Bertone, S., & Conselice, C. J. 2009, MNRAS, 396, 2345

Bertone, S., De Lucia, G., & Thomas, P. A. 2007, MNRAS, 379, 1143

Bhattacharya, D., & van den Heuvel, E. P. J. 1991, Physics Reports, 203, 1

Blanchet, L. 2014, Living Reviews in Relativity, 17, 2

Blandford, R., Romani, R. W., & Narayan, R. 1984, Journal of Astrophysics and Astronomy, 5, 369

Blandford, R. D., & Begelman, M. C. 1999, MNRAS, 303, L1

Bogdanovic, T. 2014, ArXiv e-prints, arXiv:1406.5193 [astro-ph.HE]

Bond, J. R., Arnett, W. D., & Carr, B. J. 1984, ApJ, 280, 825

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Bond, J. R., & Efstathiou, G. 1987, MNRAS, 226, 655

Bond, J. R., & Szalay, A. S. 1983, ApJ, 274, 443

Boroson, T. A., & Lauer, T. R. 2009, Nature, 458, 53

Bournaud, F., Jog, C. J., & Combes, F. 2005, A&A, 437, 69

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2014, ArXiv e-prints, arXiv:1403.4295

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

Boyle, L., & Pen, U.-L. 2012, Phys. Rev. D, 86, 124028

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23

Bromm, V., & Yoshida, N. 2011, ARA&A, 49, 373

Bromm, V., Yoshida, N., & Hernquist, L. 2003, ApJL, 596, L135

Bromm, V., Yoshida, N., Hernquist, L., & McKee, C. F. 2009, Nature, 459, 49

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Bundy, K., Fukugita, M., Ellis, R. S., et al. 2009, ApJ, 697, 1369

Burgay, M., D'Amico, N., Possenti, A., et al. 2003, Nature, 426, 531

Burke-Spolaor, S. 2013, Classical and Quantum Gravity, 30, 224013

Callegari, S., Mayer, L., Kazantzidis, S., et al. 2009, ApJ, 696, L89

Cappellari, M., Emsellem, E., Krajnović, D., et al. 2011, MNRAS, 416, 1680

Carr, B. 2003, Primordial black holes, ed. G. W. Gibbons, E. P. S. Shellard, & S. J. Rankin, 236

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Casertano, S., Phinney, E. S., & Villumsen, J. V. 1987, in IAU Symposium, Vol. 127, Structure and Dynamics of Elliptical Galaxies, ed. P. T. de Zeeuw & S. D. Tremaine, 475

Cen, R. 2011, ApJ, 741, 99

Chamberlin, S. J., & Siemens, X. 2012, Phys. Rev. D, 85, 082001

Chandrasekhar, S. 1943, ApJ, 97, 255

Chokshi, A., & Turner, E. L. 1992, MNRAS, 259, 421

Christodoulou, D. 1991, PRL, 67, 1486

Cognard, I., & Backer, D. C. 2004, ApJL, 612, L125

Cole, T. W. 1969, Nature, 221, 29

Coles, W., Hobbs, G., Champion, D. J., Manchester, R. N., & Verbiest, J. P. W. 2011, MNRAS, 418, 561

Coles, W. A., & Filice, J. P. 1984, Nature, 312, 251

Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039

Colpi, M. 2014, ArXiv e-prints, arXiv:1407.3102

Conselice, C. J. 2003, ApJS, 147, 1

—. 2014, ArXiv e-prints, arXiv:1403.2783

Conselice, C. J., Yang, C., & Bluck, A. F. L. 2009, MNRAS, 394, 1956

Corbin, V., & Cornish, N. J. 2010, ArXiv e-prints, arXiv:1008.1782 [astro-ph.HE]

Cordes, J. M., & Jenet, F. A. 2012, ApJ, 752, 54

Cordes, J. M., Kramer, M., Lazio, T. J. W., et al. 2004, New Astronomy Reviews, 48, 1413

Cordes, J. M., & Lazio, T. J. W. 2002, ArXiv Astrophysics e-prints,

astro-ph/0207156

Cordes, J. M., & Shannon, R. M. 2010, ArXiv e-prints, arXiv:1010.3785 [astro-ph.IM]

Cornish, N. J., & Sesana, A. 2013, Classical and Quantum Gravity, 30, 224005

Courteau, S., Cappellari, M., de Jong, R. S., et al. 2014, Reviews of Modern Physics, 86, 47

Cox, T. J., Jonsson, P., Somerville, R. S., Primack, J. R., & Dekel, A. 2008, MNRAS, 384, 386

Croton, D. J., Springel, V., White, S. D. M., et al. 2006, MNRAS, 365, 11

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

Davis, T. A., Bureau, M., Cappellari, M., Sarzi, M., & Blitz, L. 2013, Nature, 494, 328

De Lucia, G., & Blaizot, J. 2007, MNRAS, 375, 2

de Ravel, L., Le Fèvre, O., Tresse, L., et al. 2009, A&A, 498, 379

Deane, R. P., Paragi, Z., Jarvis, M. J., et al. 2014, Nature, 511, 57

Dehnen, W. 1993, MNRAS, 265, 250

Demorest, P. B., Ferdman, R. D., Gonzalez, M. E., et al. 2013, ApJ, 762, 94

Detweiler, S. 1979, ApJ, 234, 1100

Dicke, R. H., ed. 1970, Gravitation and the universe

Diemand, J., Moore, B., & Stadel, J. 2005, Nature, 433, 389

Dijkstra, M., Haiman, Z., Mesinger, A., & Wyithe, J. S. B. 2008, MNRAS, 391, 1961

Dijkstra, M., Haiman, Z., Rees, M. J., & Weinberg, D. H. 2004, ApJ, 601, 666

Domínguez, A., Primack, J. R., Rosario, D. J., et al. 2011, MNRAS, 410, 2556

Dotti, M., Colpi, M., & Haardt, F. 2006, MNRAS, 367, 103

Dotti, M., Sesana, A., & Decarli, R. 2012, Advances in Astronomy, 2012, arXiv:1111.0664 [astro-ph.CO]

Dressler, A., & Richstone, D. O. 1988, ApJ, 324, 701

Duffy, A. R., Wyithe, J. S. B., Mutch, S. J., & Poole, G. B. 2014, ArXiv e-prints, arXiv:1405.7459

Duquennoy, A., & Mayor, M. 1991, A&A, 248, 485

Edwards, R. T., Hobbs, G. B., & Manchester, R. N. 2006, MNRAS, 372, 1549

Efstathiou, G., Fall, S. M., & Hogan, C. 1979, MNRAS, 189, 203

Efstathiou, G., Frenk, C. S., White, S. D. M., & Davis, M. 1988, MNRAS, 235, 715

Einasto, J. 2013, Brazilian Journal of Physics, 43, 369

Einstein, A. 1916, Annalen der Physik, 354, 769

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560

Ellis, J. A., Siemens, X., & Creighton, J. D. E. 2012, ApJ, 756, 175

Emsellem, E., Cappellari, M., Krajnović, D., et al. 2011, MNRAS, 414, 888

Enoki, M., Inoue, K. T., Nagashima, M., & Sugiyama, N. 2004, ApJ, 615, 19

Enoki, M., & Nagashima, M. 2007, Progress of Theoretical Physics, 117, 241

Eracleous, M., Boroson, T. A., Halpern, J. P., & Liu, J. 2012, ApJS, 201, 23

Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. 2004, ApJ, 607, 765

—. 2005, ApJ, 630, 152

Estabrook, F. B., & Wahlquist, H. D. 1975, General Relativity and Gravitation, 6, 439

Faber, S. M., Dressler, A., Davies, R. L., Burstein, D., & Lynden-Bell, D. 1987, in Nearly Normal Galaxies. From the Planck Time to the Present, ed. S. M. Faber, 175

Faber, S. M., Tremaine, S., Ajhar, E. A., et al. 1997, AJ, 114, 1771

Fabian, A. C. 1994, ARA&A, 32, 277

—. 1999, MNRAS, 308, L39

—. 2012, ARA&A, 50, 455

Fabian, A. C., Johnstone, R. M., Sanders, J. S., et al. 2008, Nature, 454, 968

Fakhouri, O., & Ma, C.-P. 2008, MNRAS, 386, 577

Fan, L., Lapi, A., De Zotti, G., & Danese, L. 2008, ApJL, 689, L101

Fan, X. 2006, New Astronomy Reviews, 50, 665

Fan, X., Carilli, C. L., & Keating, B. 2006, ARA&A, 44, 415

Fan, X., Narayanan, V. K., Lupton, R. H., et al. 2001, AJ, 122, 2833

Fanidakis, N., Macciò, A. V., Baugh, C. M., Lacey, C. G., & Frenk, C. S. 2013, MNRAS, 436, 315

Favata, M. 2009, ApJL, 696, L159

Ferrarese, L. 2002, ApJ, 578, 90

Ferrarese, L., & Merritt, D. 2000, ApJL, 539, L9

Finkelstein, S. L., Papovich, C., Dickinson, M., et al. 2013, Nature, 502, 524

Finn, L. S., & Lommen, A. N. 2010, ApJ, 718, 1400

Foster, R. S., & Backer, D. C. 1990, ApJ, 361, 300

Frank, J., & Rees, M. J. 1976, MNRAS, 176, 633

Fryer, C. L., Woosley, S. E., & Heger, A. 2001, ApJ, 550, 372

Gaddum, J. H. 1945, Nature, 156, 3964

Gair, J. R., & Glampedakis, K. 2006, Phys. Rev. D, 73, 064037

Gebhardt, K., & Thomas, J. 2009, ApJ, 700, 1690

Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJL, 539, L13

Gebhardt, K., Richstone, D., Tremaine, S., et al. 2003, ApJ, 583, 92

Geehan, J. J., Fardal, M. A., Babul, A., & Guhathakurta, P. 2006, MNRAS, 366, 996

Genel, S., Vogelsberger, M., Springel, V., et al. 2014, ArXiv e-prints, arXiv:1405.3749

Gerosa, D., & Sesana, A. 2014, ArXiv e-prints, arXiv:1405.2072

Ghez, A. M., Salim, S., Weinberg, N. N., et al. 2008, ApJ, 689, 1044

Gingold, R. A., & Monaghan, J. J. 1977, MNRAS, 181, 375

Glampedakis, K., Hughes, S. A., & Kennefick, D. 2002, Phys. Rev. D, 66, 064005

Gold, T. 1968, Nature, 218, 731

Graham, A. W. 2012, ApJ, 746, 113

Graham, A. W., Onken, C. A., Athanassoula, E., & Combes, F. 2011, MNRAS, 412, 2211

Guo, Q., White, S., Boylan-Kolchin, M., et al. 2011, MNRAS, 413, 101

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Guth, A. H., & Pi, S.-Y. 1982, PRL, 49, 1110

Haehnelt, M. G., Natarajan, P., & Rees, M. J. 1998, MNRAS, 300, 817

Haiman, Z. 2013, in Astrophysics and Space Science Library, Vol. 396, , 293

Haiman, Z., Abel, T., & Rees, M. J. 2000, ApJ, 534, 11

Haiman, Z., & Holder, G. P. 2003, ApJ, 595, 1

Haiman, Z., & Loeb, A. 1998, ApJ, 503, 505

—. 2001, ApJ, 552, 459

Hankins, T. H., & Rickett, B. J. 1975, Methods in Computational Physics, 14, 55

Häring, N., & Rix, H.-W. 2004, ApJL, 604, L89

Harrison, E. R. 1970, Phys. Rev. D, 1, 2726

Harry, G. M., & LIGO Scientific Collaboration. 2010, Classical and Quantum Gravity, 27, 084006

Hashimoto, Y., Funato, Y., & Makino, J. 2003, ApJ, 582, 196

Haslam, C. G. T., Salter, C. J., Stoffel, H., & Wilson, W. E. 1982, A&AS, 47, 1

Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109

Hellings, R. W., & Downs, G. S. 1983, ApJL, 265, L39

Hewish, A., Bell, S. J., Pilkington, J. D. H., Scott, P. F., & Collins, R. A. 1968,

Nature, 217, 709

Ho, L. C. 2007, ApJ, 668, 94

Hobbs, G. 2013, Classical and Quantum Gravity, 30, 224007

Hobbs, G., Dai, S., Manchester, R. N., et al. 2014, ArXiv e-prints, arXiv:1407.0435 [astro-ph.IM]

Hobbs, G., Lorimer, D. R., Lyne, A. G., & Kramer, M. 2005, MNRAS, 360, 974

Hobbs, G., Lyne, A. G., & Kramer, M. 2010a, MNRAS, 402, 1027

Hobbs, G., Jenet, F., Lee, K. J., et al. 2009, MNRAS, 394, 1945

Hobbs, G., Archibald, A., Arzoumanian, Z., et al. 2010b, Classical and Quantum Gravity, 27, 084013

Hobbs, G. B., Edwards, R. T., & Manchester, R. N. 2006, MNRAS, 369, 655

Hoekstra, H., Bartelmann, M., Dahle, H., et al. 2013, Space Science Reviews, 177, 75

Hopkins, A. M., & Beacom, J. F. 2006, ApJ, 651, 142

Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731

Hopkins, P. F., Croton, D., Bundy, K., et al. 2010, ApJ, 724, 915

Hotan, A. W., Bailes, M., & Ord, S. M. 2004, MNRAS, 355, 941

Hoyle, F., & Fowler, W. A. 1963, MNRAS, 125, 169

Hubble, E. 1926a, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 324, 1

Hubble, E. P. 1925, The Observatory, 48, 139

—. 1926b, ApJ, 64, 321

—. 1936, Realm of the Nebulae

Hughes, S. A. 2002, MNRAS, 331, 805

Hulse, R. A., & Taylor, J. H. 1975, ApJL, 195, L51

Jaffe, A. H., & Backer, D. C. 2003, ApJ, 583, 616

Jenet, F. A., Hobbs, G. B., Lee, K. J., & Manchester, R. N. 2005, ApJL, 625, L123

Jenet, F. A., Lommen, A., Larson, S. L., & Wen, L. 2004, ApJ, 606, 799

Jenet, F. A., Hobbs, G. B., van Straten, W., et al. 2006, ApJ, 653, 1571

Jetsu, L., Porceddu, S., Lyytinen, J., et al. 2013, ApJ, 773, 1

Jiang, C. Y., Jing, Y. P., Faltenbacher, A., Lin, W. P., & Li, C. 2008, ApJ, 675, 1095

Jiang, L., Fan, X., Vestergaard, M., et al. 2007, AJ, 134, 1150

Johnston, S., Manchester, R. N., Lyne, A. G., et al. 1992, ApJL, 387, L37

Kaspi, V. M., Taylor, J. H., & Ryba, M. F. 1994, ApJ, 428, 713

Katz, N., Hernquist, L., & Weinberg, D. H. 1992, ApJL, 399, L109

Kauffmann, G., & Charlot, S. 1998, MNRAS, 294, 705

Kauffmann, G., & Haehnelt, M. 2000, MNRAS, 311, 576

Keith, M. J., Coles, W., Shannon, R. M., et al. 2013, MNRAS, 429, 2161

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

Khan, F. M., Just, A., & Merritt, D. 2011, ApJ, 732, 89

Khan, F. M., Preto, M., Berczik, P., et al. 2012, ApJ, 749, 147

Khandai, N., Di Matteo, T., Croft, R., et al. 2014, ArXiv e-prints, arXiv:1402.0888 [astro-ph.CO]

Khochfar, S., & Silk, J. 2006, ApJL, 648, L21

Kitayama, T., Yoshida, N., Susa, H., & Umemura, M. 2004, ApJ, 613, 631

Kitzbichler, M. G., & White, S. D. M. 2008, MNRAS, 391, 1489

Kormendy, J., & Bender, R. 2011, Nature, 469, 377

Kormendy, J., Bender, R., & Cornell, M. E. 2011, Nature, 469, 374

Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511

Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581

Kovner, I. 1989, ApJ, 337, 621

Kramer, M., & Champion, D. J. 2013, Classical and Quantum Gravity, 30, 224009

Kudryavtseva, N. A., Britzen, S., Witzel, A., et al. 2011, A&A, 526, A51

Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N., & Cen, R. 2013, ArXiv e-prints, arXiv:1307.3684 [astro-ph.CO]

Kuo, C. Y., Braatz, J. A., Condon, J. J., et al. 2011, ApJ, 727, 20

Lacey, C., & Cole, S. 1993, MNRAS, 262, 627

Lackner, C. N., & Gunn, J. E. 2012, MNRAS, 421, 2277

Lauer, T. R., Tremaine, S., Richstone, D., & Faber, S. M. 2007, ApJ, 670, 249

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Lazio, T. J. W. 2013, Classical and Quantum Gravity, 30, 224011

Lee, K. J., Jenet, F. A., & Price, R. H. 2008, ApJ, 685, 1304

Lee, K. J., Wex, N., Kramer, M., et al. 2011, MNRAS, 414, 3251

Lemson, G., & Virgo Consortium, t. 2006, ArXiv Astrophysics e-prints, astro-ph/0608019

Lentati, L., Alexander, P., Hobson, M. P., et al. 2014, MNRAS, 437, 3004

—. 2013, Phys. Rev. D, 87, 104021

Levine, J. L. 2004, Physics in Perspective, 6, 42

Li, C., & White, S. D. M. 2009, MNRAS, 398, 2177

Li, D., Nan, R., & Pan, Z. 2013, in IAU Symposium, Vol. 291, IAU Symposium, 325

Linde, A. D. 1982, Physics Letters B, 108, 389

Liu, J., Zhang, Y., Zhang, H., Sun, Y., & Wang, N. 2012, A&A, 540, A67

Liu, K., Desvignes, G., Cognard, I., et al. 2014, ArXiv e-prints, arXiv:1407.3827

Lodato, G., & Natarajan, P. 2006, MNRAS, 371, 1813

Łokas, E. L., & Mamon, G. A. 2001, MNRAS, 321, 155

Lommen, A. N. 2002, in Neutron Stars, Pulsars, and Supernova Remnants, ed. W. Becker, H. Lesch, & J. Trümper, 114

López-Sanjuan, C., Le Fèvre, O., & Ilbert, O. e. a. 2012, A&A, 548, A7

Lorimer, D. R., & Kramer, M. 2012, Handbook of Pulsar Astronomy

Lotz, J. M., Jonsson, P., Cox, T. J., & Primack, J. R. 2008, MNRAS, 391, 1137

—. 2010, MNRAS, 404, 575

Lotz, J. M., Papovich, C., Faber, S. M., et al. 2013, ApJ, 773, 154

Lynden-Bell, D. 1967, MNRAS, 136, 101

—. 1969, Nature, 223, 690

Lynden-Bell, D., & Rees, M. J. 1971, MNRAS, 152, 461

Lyne, A. G., Burgay, M., Kramer, M., et al. 2004, Science, 303, 1153

Lyth, D. H. D. H., & Riotto, A. A. 1999, Physics Reports, 314, 1

Madau, P., & Dickinson, M. 2014, ArXiv e-prints, arXiv:1403.0007

Madison, D. R., Cordes, J. M., & Chatterjee, S. 2014, ApJ, 788, 141

Maggiore, M. 2000, ArXiv General Relativity and Quantum Cosmology e-prints, gr-qc/0008027

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Manchester, R. N., Hobbs, G. B., Teoh, A., & Hobbs, M. 2005, AJ, 129, 1993

Manchester, R. N., & IPTA. 2013, Classical and Quantum Gravity, 30, 224010

Manchester, R. N., Hobbs, G., Bailes, M., et al. 2013, PASA, 30, 17

Marconi, A., & Hunt, L. K. 2003, ApJL, 589, L21

Marulli, F., Bonoli, S., Branchini, E., Moscardini, L., & Springel, V. 2008, MNRAS, 385, 1846

Massey, R., Kitching, T., & Richard, J. 2010, Reports on Progress in Physics, 73, 086901

Matthews, T. A., & Sandage, A. R. 1963, ApJ, 138, 30

Mayer, L., Kazantzidis, S., Madau, P., et al. 2007, Science, 316, 1874

McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

McConnell, N. J., Ma, C.-P., Gebhardt, K., et al. 2011, Nature, 480, 215

McKee, C. F., & Tan, J. C. 2008, ApJ, 681, 771

McLaughlin, M. A. 2013, Classical and Quantum Gravity, 30, 224008

McLure, R. J., & Dunlop, J. S. 2002, MNRAS, 331, 795

McNamara, B. R., & Nulsen, P. E. J. 2012, New Journal of Physics, 14, 055023

McWilliams, S. T., Ostriker, J. P., & Pretorius, F. 2014, ApJ, 789, 156

Meert, A., Vikram, V., & Bernardi, M. 2014, ArXiv e-prints, arXiv:1406.4179

Mendel, J. T., Simard, L., Palmer, M., Ellison, S. L., & Patton, D. R. 2014, ApJS, 210, 3

Merritt, D. 2001, ApJ, 556, 245

Merritt, D., & Milosavljević, M. 2005, Living Reviews in Relativity, 8, 8

Merritt, D., & Poon, M. Y. 2004, ApJ, 606, 788

Miller, B. P., Gallo, E., Greene, J. E., et al. 2014, ArXiv e-prints, arXiv:1403.4246 [astro-ph.GA]

Milosavljević, M., & Merritt, D. 2001, ApJ, 563, 34

Milosavljević, M., & Merritt, D. 2003, in AIP Conference Series, Vol. 686, , 201

Mingarelli, C. M. F., Sidery, T., Mandel, I., & Vecchio, A. 2013, Phys. Rev. D, 88, 062005

Mitchell, P. D., Lacey, C. G., Baugh, C. M., & Cole, S. 2013, MNRAS, 435, 87

Moore, C. J., Taylor, S. R., & Gair, J. R. 2014, ArXiv e-prints, arXiv:1406.5199 [astro-ph.IM]

Mortlock, D. J., Warren, S. J., Venemans, B. P., et al. 2011, Nature, 474, 616

Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903

Mutch, S. J., Poole, G. B., & Croton, D. J. 2013, MNRAS, 428, 2001

Muzzin, A., Marchesini, D., Stefanon, M., et al. 2013, ApJ, 777, 18

Narlikar, J. V., & Padmanabhan, T. 1991, ARA&A, 29, 325

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, ApJ, 786, 108

Oh, S. P., & Haiman, Z. 2002, ApJ, 569, 558

Omukai, K., Schneider, R., & Haiman, Z. 2008, ApJ, 686, 801

Opik, E. 1922, ApJ, 55, 406

Oppenheimer, J. R., & Volkoff, G. M. 1939, Phys. Rev., 55, 374

Osłowski, S., van Straten, W., Demorest, P., & Bailes, M. 2013, MNRAS, 430, 416

Osłowski, S., van Straten, W., Hobbs, G. B., Bailes, M., & Demorest, P. 2011, MNRAS, 418, 1258

Ostriker, J. P., & Peebles, P. J. E. 1973, Ap, 186, 467

Ostriker, J. P., Peebles, P. J. E., & Yahil, A. 1974, ApJL, 193, L1

Pacini, F. 1967, Nature, 216, 567

Paczynski, B. 1986, ApJ, 304, 1

Page, T. 1960, ApJ, 132, 910

Papitto, A., Ferrigno, C., Bozzo, E., et al. 2013, Nature, 501, 517

Peebles, P. J. E. 1982, ApJL, 263, L1

Peebles, P. J. E., & Yu, J. T. 1970, ApJ, 162, 815

Peng, C. Y. 2007, ApJ, 671, 1098

Pennucci, T. T., Demorest, P. B., & Ransom, S. M. 2014, ArXiv e-prints, arXiv:1402.1672 [astro-ph.IM]

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Peters, P. C., & Mathews, J. 1963, Phys. Rev., 131, 435

Peterson, B. M. 1993, PASP, 105, 247

Peterson, J. R., & Fabian, A. C. 2006, Physics Reports, 427, 1

Phinney, E. S. 2001, ArXiv Astrophysics e-prints, astro-ph/0108028

Pitkin, M. 2012, MNRAS, 425, 2688

Pitkin, M., Reid, S., Rowan, S., & Hough, J. 2011, Living Reviews in Relativity, 14, 5

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013, ArXiv e-prints, arXiv:1303.5076 [astro-ph.CO]

Polnarev, A. G., & Rees, M. J. 1994, A&A, 283, 301

Pounds, K. 2013, Space Science Reviews

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Preto, M., Berentzen, I., Berczik, P., Merritt, D., & Spurzem, R. 2009, Journal of Physics Conference Series, 154, 012049

Preto, M., Berentzen, I., Berczik, P., & Spurzem, R. 2011, ApJL, 732, L26

Quinlan, G. D. 1996, New Astronomy, 1, 35

Quinlan, G. D., & Hernquist, L. 1997, New Astronomy, 2, 533

Quinlan, G. D., & Shapiro, S. L. 1990, ApJ, 356, 483

Radhakrishnan, V., Cooke, D. J., Komesaroff, M. M., & Morris, D. 1969, Nature, 221, 443

Radhakrishnan, V., & Manchester, R. N. 1969, Nature, 222, 228

Rajagopal, M., & Romani, R. W. 1995, ApJ, 446, 543

Ravi, V., Wyithe, J. S. B., Hobbs, G., et al. 2012, ApJ, 761, 84

Ravi, V., Wyithe, J. S. B., Shannon, R. M., & Hobbs, G. 2015, MNRAS, 447, 2772

Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. 2014, MNRAS, 442, 56

Rees, M. J., & Ostriker, J. P. 1977, MNRAS, 179, 541

Ricotti, M., Gnedin, N. Y., & Shull, J. M. 2001, ApJ, 560, 580

—. 2002, ApJ, 575, 49

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Roberts, M. S. 1976, Comments on Astrophysics, 6, 105

Roberts, M. S., & Whitehurst, R. N. 1975, ApJ, 201, 327

Robertson, B., Cox, T. J., Hernquist, L., et al. 2006, ApJ, 641, 21

Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71

Robotham, A. S. G., Driver, S. P., Davies, L. J. M., et al. 2014, ArXiv e-prints, arXiv:1408.1476

Rodriguez, C., Taylor, G. B., Zavala, R. T., et al. 2006, ApJ, 646, 49

Roedig, C., Dotti, M., Sesana, A., Cuadra, J., & Colpi, M. 2011, MNRAS, 415, 3033

Roedig, C., & Sesana, A. 2012, Journal of Physics Conference Series, 363, 012035

Romani, R. W., & Taylor, J. H. 1983, ApJL, 265, L35

Roškar, R., Mayer, L., Fiacconi, D., et al. 2014, ArXiv e-prints, arXiv:1406.4505

Rubin, V. C., & Ford, Jr., W. K. 1970, ApJ, 159, 379

Rusli, S. P., Thomas, J., Saglia, R. P., et al. 2013, AJ, 146, 45

Salpeter, E. E. 1964, ApJ, 140, 796

Sanders, D. B., & Mirabel, I. F. 1996, ARA&A, 34, 749

Saslaw, W. C., Valtonen, M. J., & Aarseth, S. J. 1974, ApJ, 190, 253

Sathyaprakash, B. S., & Schutz, B. F. 2009, Living Reviews in Relativity, 12, 2

Sazhin, M. V. 1978, Soviet Astronomy, 22, 36

Schechter, P. 1976, ApJ, 203, 297

Schmidt, M. 1963, Nature, 197, 1040

Schnittman, J. D. 2011, Classical and Quantum Gravity, 28, 094021

Schuster, A. 1898, Terrestrial Magnetism (Journal of Geophysical Research), 3, 13

Schwarzschild, K. 1916, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 1916, Seite 189-196, 189

Scott, N., Graham, A. W., & Schombert, J. 2013, ApJ, 768, 76

Sesana, A. 2010, ApJ, 719, 851

—. 2013a, ArXiv e-prints, arXiv:1307.2600 [astro-ph.CO]

—. 2013b, MNRAS, 433, L1

Sesana, A., Haardt, F., & Madau, P. 2006, ApJ, 651, 392

—. 2008a, ApJ, 686, 432

Sesana, A., Haardt, F., Madau, P., & Volonteri, M. 2004, ApJ, 611, 623

Sesana, A., Vecchio, A., & Colacino, C. N. 2008b, MNRAS, 390, 192

Sesana, A., Vecchio, A., & Volonteri, M. 2009, MNRAS, 394, 2255

Sethi, S., Haiman, Z., & Pandey, K. 2010, ApJ, 721, 615

Shang, C., Bryan, G. L., & Haiman, Z. 2010, MNRAS, 402, 1249

Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009, ApJ, 690, 20

—. 2013, MNRAS, 428, 421

Shannon, R. M., & Cordes, J. M. 2010, ApJ, 725, 1607

Shannon, R. M., Ravi, V., Coles, W. A., et al. 2013, Science, 342, 334

Shannon, R. M., Osłowski, S., Dai, S., et al. 2014, ArXiv e-prints, arXiv:1406.4716 [astro-ph.SR]

Shapiro, S. L. 2004, ApJ, 610, 913

Shen, Y., & Kelly, B. C. 2010, ApJ, 713, 41

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Siemens, X., Ellis, J., Jenet, F., & Romano, J. D. 2013, ArXiv e-prints, arXiv:1305.3196 [astro-ph.IM]

Silk, J. 1991, Science, 251, 537

Silk, J., & Rees, M. J. 1998, A&A, 331, L1

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11

Small, T. A., & Blandford, R. D. 1992, MNRAS, 259, 725

Smith, S. 1936, ApJ, 83, 23

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJL, 396, L1

Soltan, A. 1982, MNRAS, 200, 115

Somerville, R. S., & Kolatt, T. S. 1999, MNRAS, 305, 1

Spaans, M., & Silk, J. 2006, ApJ, 652, 902

Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175

Spoon, H. W. W., Farrah, D., Lebouteiller, V., et al. 2013, ApJ, 775, 127

Springel, V. 2010, MNRAS, 401, 791

Springel, V., White, S. D. M., Tormen, G., & Kauffmann, G. 2001, MNRAS, 328, 726

Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Nature, 435, 629

Staelin, D. H., & Reifenstein, III, E. C. 1968, Science, 162, 1481

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJL, 462, L17

Stinebring, D. R., Ryba, M. F., Taylor, J. H., & Romani, R. W. 1990, PRL, 65, 285

Sudou, H., Iguchi, S., Murata, Y., & Taniguchi, Y. 2003, Science, 300, 1263

Tanaka, T., & Haiman, Z. 2009, ApJ, 696, 1798

Tanaka, T. L. 2013, MNRAS, 434, 2275

—. 2014, ArXiv e-prints, arXiv:1406.3023

Tanaka, T. L., & Li, M. 2014, MNRAS, 439, 1092

Taylor, J. H. 1992, Royal Society of London Phil. Trans. Series A, 341, 117

Taylor, S. R., & Gair, J. R. 2013, Phys. Rev. D, 88, 084001

Tescari, E., Katsianis, A., Wyithe, J. S. B., et al. 2014, MNRAS, 438, 3490

Thorne, K. S. 1980, Reviews of Modern Physics, 52, 299

Toomre, A., & Toomre, J. 1972, ApJ, 178, 623

Treister, E., Schawinski, K., Volonteri, M., & Natarajan, P. 2013, ApJ, 778, 130

Turner, E. L. 1991, AJ, 101, 5

Ulrich, M.-H., Maraschi, L., & Urry, C. M. 1997, ARA&A, 35, 445

Valtonen, M. J., Lehto, H. J., Nilsson, K., et al. 2008, Nature, 452, 851

van den Bergh, S. 1962, Zeitschrift f§r Astrophysik, 55, 21

—. 1999, PASP, 111, 657

van den Bosch, F. C., Lewis, G. F., Lake, G., & Stadel, J. 1999, ApJ, 515, 50

van Haasteren, R., & Levin, Y. 2010, MNRAS, 401, 2372

van Haasteren, R., Levin, Y., McDonald, P., & Lu, T. 2009, MNRAS, 395, 1005

van Haasteren, R., Levin, Y., Janssen, G. H., et al. 2011, MNRAS, 414, 3117

van Straten, W. 2006, ApJ, 642, 1004

Van Waerbeke, L., Mellier, Y., Pelló, R., et al. 2002, A&A, 393, 369

Van Wassenhove, S., Volonteri, M., Mayer, L., et al. 2012, ApJL, 748, L7

Vasiliev, E., Antonini, F., & Merritt, D. 2014, ApJ, 785, 163

Verbiest, J. P. W. 2008, PhD thesis, Ph. D. thesis, Swinburne University of Technology (2008).

Verbiest, J. P. W., Bailes, M., Coles, W. A., et al. 2009, MNRAS, 400, 951

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, Nature, 509, 177

Volonteri, M. 2010, A&Ar, 18, 279

Volonteri, M., & Bellovary, J. 2012, Reports on Progress in Physics, 75, 124901

Volonteri, M., Haardt, F., & Madau, P. 2003, ApJ, 582, 559

Volonteri, M., & Rees, M. J. 2005, ApJ, 633, 624

Wahlquist, H. 1987, General Relativity and Gravitation, 19, 1101

Wang, J. B., Hobbs, G., Coles, W., et al. 2015, MNRAS, 446, 1657

Warszawski, L., & Melatos, A. 2008, MNRAS, 390, 175

Weber, J. 1967, PRL, 18, 498

—. 1969, PRL, 22, 1320

White, S. D. M., Davis, M., & Frenk, C. S. 1984, MNRAS, 209, 27P

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52

White, S. D. M., Frenk, C. S., & Davis, M. 1983, ApJL, 274, L1

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Will, C. M. 2014, Living Reviews in Relativity, 17, 4

Willott, C. J. 2011, ApJL, 742, L8

Willott, C. J., Delorme, P., Reylé, C., et al. 2010, AJ, 139, 906

Wise, J. H., & Abel, T. 2007, ApJ, 665, 899

Wolszczan, A., & Frail, D. A. 1992, Nature, 355, 145

Wyithe, J. S. B., & Loeb, A. 2003a, ApJ, 590, 691

—. 2003b, ApJ, 595, 614

—. 2012, MNRAS, 425, 2892

Xu, C. K., Zhao, Y., Scoville, N., et al. 2012, ApJ, 747, 85

Yardley, D. R. B., Hobbs, G. B., Jenet, F. A., et al. 2010, MNRAS, 407, 669

Yardley, D. R. B., Coles, W. A., Hobbs, G. B., et al. 2011, MNRAS, 414, 1777

Yu, Q. 2002, MNRAS, 331, 935

Yu, Q., Lu, Y., Mohayaee, R., & Colin, J. 2011, ApJ, 738, 92

Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965

Zel'dovich, Y. B. 1964, Soviet Physics Doklady, 9, 195

—. 1972, MNRAS, 160, 1P

Zhang, W., Woosley, S. E., & Heger, A. 2008, ApJ, 679, 639

Zhu, X.-J., Hobbs, G., Wen, L., et al. 2014, MNRAS, 444, 3709

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

—. 1937, ApJ, 86, 217