

HARVARD UNIVERSITY  
Graduate School of Arts and Sciences



**DISSERTATION ACCEPTANCE CERTIFICATE**

The undersigned, appointed by the

**Department of Astronomy**

have examined a dissertation entitled

***Massive Black-Hole Binary Mergers: Dynamics, Environments & Expected Detections***

presented by Luke Zoltan Kelley

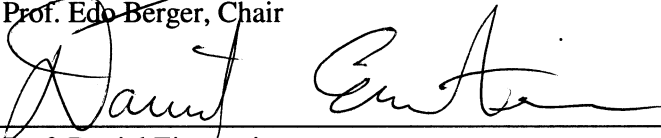
candidate for the degree of Doctor of Philosophy, and hereby

certify that it is worthy of acceptance.

Signature

  
Typed name: Prof. Edo Berger, Chair

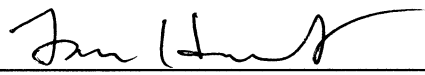
Signature

  
Typed name: Prof. Daniel Eisenstein

Signature

  
Typed name: Prof. Zoltan Haiman

Signature

  
Typed name: Prof. Lars Hernquist

Date: April 13, 2018



# Massive Black-Hole Binary Mergers: Dynamics, Environments & Expected Detections

A DISSERTATION PRESENTED  
BY  
LUKE ZOLTAN KELLEY  
TO  
THE DEPARTMENT OF ASTRONOMY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY  
IN THE SUBJECT OF  
ASTRONOMY AND ASTROPHYSICS

HARVARD UNIVERSITY  
CAMBRIDGE, MASSACHUSETTS  
APRIL 2018

©2018 – LUKE ZOLTAN KELLEY  
ALL RIGHTS RESERVED.

# Massive Black-Hole Binary Mergers: Dynamics, Environments & Expected Detections

## ABSTRACT

This thesis studies the populations and dynamics of massive black-hole binaries and their mergers, and explores the implications for electromagnetic and gravitational-wave signals that will be detected in the near future. Massive black-holes (MBH) reside in the centers of galaxies, and when galaxies merge, their MBH interact and often pair together. We base our study on the populations of MBH and galaxies from the ‘Illustris’ cosmological hydrodynamic simulations. The bulk of the binary merger dynamics, however, are unresolved in cosmological simulations. We implement a suite of comprehensive physical models for the merger process, like dynamical friction and gravitational wave emission, which are added in post-processing. Contrary to many previous studies, we find that the most massive binaries with near equal-mass companions are the most efficient at coalescing; though the process still typically takes gigayears.

From the data produced by these MBH binary populations and their dynamics, we calculate the expected gravitational wave (GW) signals: both the stochastic, GW background of countless unresolved sources, and the GW foreground of individually resolvable binaries which resound above the noise. Ongoing experiments, called pulsar timing arrays, are sensitive to both of these types of signals. We find that, while the current lack of detections is unsurprising, both the background and foreground will plausibly be detected in the next decade. Unlike previous studies which have predicted the foreground to be significantly harder to detect than the background, we find their typical amplitudes are comparable.

With traditional electromagnetic observations, there has also been a dearth of confirmed detections of MBH binary systems. We use our binaries, combined with models of emission from accreting MBH systems, to make predictions for the occurrence rate of systems observable using photometric, periodic-variability surveys. These variables should be detectable in current surveys, and indeed, we expect many candidates recently identified to be true binaries - though a significant fraction are likely false positives. Overall, this thesis finds the science of MBH binaries at an exciting cusp: just before incredible breakthroughs in observations, both electromagnetically and in the new age of gravitational wave astrophysics.

# CONTENTS

|       |  |     |
|-------|--|-----|
| 0     | INTRODUCTION   | 1   |
| 0.1   | Massive Black Holes and Galaxies   | 2   |
| 0.1.1 | Cosmological Simulations   | 3   |
| 0.1.2 | Illustris  | 5   |
| 0.1.3 | Binary Dynamics  | 6   |
| 0.2   | Gravitational Wave Signals & Pulsar Timing Arrays  | 8   |
| 0.3   | Electromagnetic Observations   | 10  |
| 0.4   | Thesis Preview   | 11  |
| 1     | MASSIVE BLACK HOLE BINARY MERGERS IN DYNAMICAL GALACTIC ENVIRONMENTS   | 13  |
| 1.1   | Introduction   | 13  |
| 1.2   | The Illustris Simulations  | 20  |
| 1.2.1 | The Black Hole Merger Population   | 21  |
| 1.2.2 | Merger Host Galaxies   | 22  |
| 1.3   | Binary Hardening Models  | 24  |
| 1.3.1 | Dynamical Friction   | 24  |
| 1.3.2 | Stellar Loss-Cone Scattering   | 28  |
| 1.3.3 | Viscous Hardening by a Circumbinary Disk   | 30  |
| 1.3.4 | Gravitational-Wave Emission  | 37  |
| 1.4   | Results  | 37  |
| 1.4.1 | Binary Lifetimes   | 38  |
| 1.4.2 | The Gravitational Wave Background (GWB)  | 44  |
| 1.4.3 | The Populations of MBHB  | 53  |
| 1.4.4 | MBH Triples  | 58  |
| 1.5   | Conclusions and Summary  | 59  |
| A     | Supplemental Material  | 63  |
| 2     | THE GRAVITATIONAL WAVE BACKGROUND FROM MASSIVE BLACK HOLE BINARIES IN ILLUSTRIS: SPECTRAL FEATURES AND TIME TO DETECTION WITH PULSAR TIMING ARRAYS | 71  |
| 2.1   | Introduction   | 72  |
| 2.2   | Methods  | 73  |
| 2.2.1 | Illustris MBH Mergers and Environments   | 73  |
| 2.2.2 | Models for Eccentric Binary Evolution  | 73  |
| 2.2.3 | Gravitational Waves from Eccentric MBH Binaries  | 75  |
| 2.2.4 | Detection with Pulsar Timing Arrays  | 77  |
| 2.3   | Results  | 80  |
| 2.3.1 | Eccentric Evolution  | 80  |
| 2.3.2 | Gravitational Wave Backgrounds   | 84  |
| 2.3.3 | Pulsar Timing Array Detections   | 88  |
| 2.4   | Conclusions  | 93  |
| 3     | SINGLE SOURCES IN THE LOW-FREQUENCY GRAVITATIONAL WAVE SKY: PROPERTIES AND TIME TO DETECTION BY PULSAR TIMING ARRAYS                               | 106 |
| 3.1   | Introduction   | 106 |
| 3.2   | Methods  | 109 |
| 3.2.1 | MBH Binary Population and Evolution  | 109 |
| 3.2.2 | Detection Statistics and Pulsar Timing Array Models  | 110 |
| 3.3   | Results  | 112 |
| 3.3.1 | The Structure of the Low-Frequency GW Sky  | 112 |

|          |   |            |
|----------|---|------------|
| 3.3.2    | Parametric Dependencies                                 | 117        |
| 3.3.3    | Times to Detection                                      | 120        |
| 3.4      | Discussion  | 125        |
| 3.4.1    | Caveats   | 125        |
| 3.4.2    | Conclusions   | 125        |
| <b>4</b> | <b>MASSIVE BH BINARIES AS PERIODICALLY-VARIABLE AGN</b> | <b>134</b> |
| 4.1      | Introduction  | 137        |
| 4.2      | Methods   | 138        |
| 4.2.1    | MBH Binary Population and Evolution                     | 138        |
| 4.2.2    | MBH Accretion and AGN Spectra                           | 139        |
| 4.2.3    | Observation Parameters and Calculations                 | 142        |
| 4.2.4    | Models of Variability                                   | 143        |
| 4.3      | Results   | 144        |
| 4.3.1    | Doppler Variable Binaries                               | 144        |
| 4.3.2    | Hydrodynamically Variable Binaries                      | 148        |
| 4.4      | Discussion  | 150        |
| 4.4.1    | Caveats   | 150        |
| 4.4.2    | Conclusions   | 151        |
| <b>5</b> | <b>CONCLUSIONS</b>                                      | <b>161</b> |
| 5.1      | Summary   | 161        |
| 5.2      | The Future  | 163        |
| <b>A</b> | <b>APPENDIX</b>   | <b>165</b> |
| 1.1      | Frequently used Abbreviations                           | 165        |
| 1.2      | Frequently used Symbols and Values                      | 166        |
|          | <b>REFERENCES</b>                                       | <b>187</b> |

EURYDICE HAD NO MORE ADORATION,  
ALEXANDER NO BETTER INSTRUCTION,  
ADAM NO BETTER CREATION.

TO MOM & DAD, THE BEST TEACHERS ANYONE HAS EVER HAD,  
AND THE BEST PARENTS IN THE UNIVERSE—WHICH I NOW SAY WITH AUTHORITY.



# ACKNOWLEDGMENTS

Throughout this work I use ‘we’. Each time I am reminded of the countless coauthors, advisors, colleagues & friends who have made this thesis, and indeed my learning and career as a whole, not only possible but productive & enjoyable. Acknowledgments, even the length of this thesis, would still be woefully inadequate to describe the gracious gifts of so many people who have supported me. Thus, especially as a scientist and not a poet, mere pages—and truly, any mere words—of thanks and recognition seem insulting to such kindness and good fortune. Please take the acknowledgements below, despite their inevitable omissions, as a brief glimpse into the true depths of my gratitude. I can only say that beneath these words are a full heart, and a will determined to manifest from my great privilege, great good—to the world, to society, and to science...

First and foremost I want to acknowledge the tremendous support and encouragement provided by my advisor, Lars Hernquist. When I approached Lars four years ago, the topics I had in mind were well outside of the standard wheel-well of the prodigious and prolific Hernquist group, yet Lars was unhesitant in entertaining my ideas and welcoming me into the fold. The topics I have chosen to work on have little overlap, besides in tool sets, those that Lars has studied, but he has been no less available for discussion, advice, and always volunteers kind words which are ever important for a graduate student’s at-times-delicate morale. Lars has given me every freedom and independence to explore my own ideas and directions, and yet has always gone to great lengths to accommodate my needs. I am incredibly grateful for the opportunity to work with him, for the wisdom he has shared, and the support he has provided.

Early on, after joining the Hernquist-group, Lars introduced me to Laura Blecha—at the time a post-doc at Maryland, and now a professor at the University of Florida. Laura has truly been a second thesis advisor, throughout. The wisdom and insight in her advice and knowledge have entirely belied that she had only graduated from Harvard one or two years before we began collaborating. Working with Laura, and learning about gravitational waves together, has been as engaging and enjoyable as it has been fruitful, and I’m very excited to continue working together throughout my career.

I have already said that my research topics were outside of the normal bounds of my research group. More still, no one at Harvard or the Center for Astrophysics had worked much on these topics. I was thus incredibly fortunate to have been guided and aided along in my studies by many experts around the world without whom I surely would have been stuck, or worse—yet, completely off track. I am especially grateful to have had the guidance and opportunity to collaborate with Alberto Sesana, Alexander Tchekhovskoy, Ilya Mandel, Stephen Taylor & Zoltan Haiman. The expertise of these scientists is truly remarkable, and only paralleled by their willingness to share their knowledge. Many times, just being witness to the great discussions between these collaborators have provided incredible insights into the thought processes which underly ‘physical intuition,’ and the often mystical methodologies leading to great scientific conclusions. Importantly, no question has ever been too complex or too simple to warrant thorough consideration and discussion with any of them, no matter how busy they become.

In addition to expert collaborators, it has been a pleasure to be part of the NANOGrav community—which has been incredibly welcoming to a wandering grad-student such as myself. Participating in this collaboration has been fun, and extremely beneficial to my science. I am especially thankful for the encouragement, kindness, and excellent leadership of Maura McLaughlin. The Hernquist group, and the ex-

tended Arepo/Illustris collaboration is truly a community unto its own. The post-docs and grad-students in the group have been incredible resources for understanding everything from day-to-day code bugs, to the big picture scientific goals. While we have only minimally interacted directly, I am extremely thankful for Volker Springel's approach to science: creating and sharing the most powerful tools available, that have led to now-dozens of theses projects. If my own research has even a small fraction the impact of Volker's, I could consider my career an immense success.

The most immediate community I have been a part of has been the Harvard Astronomy & Astrophysics department. I decided to come to Harvard in large part because of the incredible graduate students here. Besides being brilliant scientists, I found people with an incredible diversity of backgrounds, interests, and passions composing a brilliant family of explorers—of life as of astronomy. For a list of the particular people who have positively impacted me, count everyone who has passed through the department between 2011 and 2018. I want to especially acknowledge my incredible friends: Elizabeth Newton, Jieun Choi, Katie Auchettl, Laura Schaefer, Maria Drout, Ryan Loomis and Sarah Wellons. Our time together has been precious, and I can only hope distance will not separate us. Similarly, the friends I have made through and outside the department, particularly, Amy Relihan, Bill Sowerwine, Mark Wilson, and Zavisla Dogic, have helped me in (the non-trivial task of) staying sane throughout these years. I look forward to many further adventures across the globe.

I will never be able to thank, repay, or even understand the unending support, encouragement and enthusiasm that Enrico Ramirez-Ruiz has shared with me. Throughout the now ten years that I have known him, Enrico has been an incredible mentor and friend, and many times has seemed to presciently pave the very path I follow—yet never deciding its direction for me. Enrico viewpoints and intuition are remarkable, and it feels like every thought he shares bestows as much physical insight as any published work. I am forever grateful.

My oldest friends, almost from a past life, Patrick Tyler Haas and Samuel Schwemberger, have been an unwavering source of perspective, empathy, and joy for so many years. As knowledgeable as my colleagues and professors are in science and astrophysics, so are Patrick and Sam in seemingly all other things. It seems all too appropriate that while surrounded by ivy-league academics, it's my buddies from high school who so easily (and usually indirectly) keep my head of humble size. They have certainly become brothers. The most important part of my time at Harvard has been meeting Jennifer Bergner, my partner. While friends have become family, Jenny has become part of myself. She is the most incredible person I have ever met. In addition to all ends of emotional and intellectual support, she has filled my life with happiness.

In acknowledging my siblings: Hillary & Mhat Bernstein and Mimi Shum, and especially my parents: Kathy & Guy Kelley, words fail me the most. The love, kindness and experiences that my siblings have shared with me are unparalleled. Their presence and their world-views have been pillars of my own perception. Besides our lives together, we are united in sharing our incredible parents. My parents are my idols; the most brilliant, reasoned, passionate and moral people I have ever known. Nothing can express my gratitude, and sheer awe, at having so incredible a family and so fortunate a life.



# INTRODUCTION

Black holes (BH) are the most exotic, unfathomable objects in the universe, yet they are virtually ubiquitous as Massive Black Holes (MBH), where they occupy the majority of massive galaxies in the universe (e.g. [Soltan 1982](#); [Kormendy & Richstone 1995](#); [Magorrian et al. 1998](#)). Especially as Active Galactic Nuclei (AGN)—the collective term for MBH producing electromagnetic radiation (e.g. [Urry & Padovani 1995](#))—MBH, with masses  $M \gtrsim 10^5 M_{\odot}$ , are also some of the most studied objects in astrophysics. Our understanding of MBH and AGN has improved dramatically over the last few decades, largely due to revolutions in computational techniques scales of both individual black-holes and their accretion disks (e.g. [Balbus & Hawley 1998](#)) and on cosmological scales of many galaxies over billions of years (e.g. [Di Matteo, Springel & Hernquist 2005](#)).

MBH Binaries (MBHB), formed after the merger of galaxies each hosting individual MBH, are of rapidly growing interest in astronomy as our understanding of galaxy evolution has reached maturity (e.g. [Binney & Tremaine 1987](#)), and the field of Gravitational Wave (GW) astrophysics has formally emerged with the detection of GW150914 by the Laser Interferometer Gravitational-Wave Observatory (LIGO; [The LIGO Scientific Collaboration & the Virgo Collaboration 2016](#)). Like their small, stellar-mass\* siblings detected by LIGO, MBHB are also expected to produce GW—but with amplitude millions of times larger. The standard measure of GW amplitude is ‘strain’: a dimensionless measure which is used to refer to two separate (but closely related) quantities. Formally, the *metric-strain* is the non-Minkowskian, spacetime-metric perturbation in linearized gravity<sup>†</sup>, i.e.  $h_{\mu\nu} \equiv g_{\mu\nu} - \eta_{\mu\nu}$ , while the phenomenological *strain-amplitude* (e.g. [Anholm et al. 2009](#)) is the observed redshift or contraction, i.e.  $h \sim \Delta t/t \sim \Delta L/L$ .

The characteristic amplitude of GW from MBH are near  $h \sim 10^{-15}$ , and typically come from the most massive systems with  $M \sim 10^9-10^{10} M_{\odot}$ . Stellar-mass objects produce strains at much smaller amplitudes,  $h \sim 10^{-22}$ , equivalent to a change in distance of a nanometer across the distance from the Earth to the Sun, making LIGO’s detections all the more impressive. In addition to GW, MBHB are also believed to play an important role in numerous other astrophysical phenomena, including: stellar tidal-disruption events ([Ivanov, Polnarev & Saha 2005a](#)), helical structures in relativistic jets ([Roos, Kaastra & Hummel 1993](#)), AGN variability ([Farris et al. 2014](#)), and MBH wandering through their hosts after being ‘kicked’ by GW recoils ([Blecha et al. 2016](#)). In this thesis, in addition to their general dynamics and evolution, we are focused on direct GW detections, and AGN variability signatures; these topics we focus on henceforth.

LIGO uses interference between orthogonal, kilometer-scale lasers to precisely measure deviations in path length, with a sensitivity of  $\sim 10^{-23}$  at frequencies  $\sim 30 - 10^3$  Hz. In this regime, LIGO is primarily sensitive to the final coalescence of roughly stellar-mass binaries (particular neutron stars and stellar-mass

\*The BH so-far detected by LIGO typically have masses of a few dozen solar-masses, but are still generally referred to as ‘stellar-mass’ BH because of their likely origin from a near-unity number of stars.

<sup>†</sup>Here, it suffices to say that  $g_{\mu\nu}$  is a description of the full, dynamic spacetime, while  $\eta_{\mu\nu}$  is the flat, static portion.

BH). The Laser Interferometer Space Antenna (LISA; [Amaro-Seoane et al. 2012](#)) is a proposed space-based interferometer with a recently successful pathfinder mission ([Armano et al. 2016](#)). Based on the same techniques as LIGO, LISA would include three, much longer baselines of  $\sim 10^6$  km, producing lower frequency sensitivities of  $\sim 10^{-2}$  Hz at amplitudes  $h \sim 10^{-20}$ . LISA will be sensitive to pre-coalescence stellar-mass binaries (including galactic white-dwarf systems) and the inspiral of higher-mass systems including intermediate-mass and massive BH ( $\sim 10^3 - 10^7 M_\odot$ ; e.g. [Amaro-Seoane et al. 2007](#)). Most MBH binaries, with  $M \gtrsim 10^7 M_\odot$  coalesce below the sensitive band of LISA, though they produce strain amplitudes drastically higher than those expected from LIGO and LISA sources, on the order of  $h \sim 10^{-15}$ . Pulsar Timing Arrays (PTA) are already reaching these sensitivities by precisely timing galactic millisecond pulsars (e.g. [Foster & Backer 1990](#)), and correlating those signals across the sky to look for patterns unique to GW signals ([Hellings & Downs 1983](#)).

Before MBH reach the small separations at which their GW signals may become observable ( $\sim 10^{-3}$  pc), they need to first come together in a galaxy merger ( $\sim 10^3$  pc), and cross the orders of magnitude in separation between those scales. This ‘hardening\*’ process is believed to be governed by interactions between the MBH and the galaxies’s dark-matter, stellar, and gaseous components. Electromagnetic observations of AGN in merging galaxies, and systems with AGN exhibiting signatures of multiplicity, are beginning to provide insights and constraints into our understanding of this exotic and extreme evolution. Excitingly, we are just reaching a frontier in which PTA constraints on GW, telescopic observations of AGN, and cosmological simulations of MBH binaries are all coming together. Together, the evidence suggest that we are on the precipice of making measurements from sup-parsec separation MBHB identified electromagnetically, and also detecting GW signals from these incredible, inspiraling behemoths.

In the remainder of this chapter, we present an introduction to the topics underlying this thesis. In §0.1 we introduce massive black holes and their galactic hosts, with a discussion of the numerical methods used to model these systems in cosmological environments (§0.1.1), focusing on the Illustris simulations which we use extensively (§0.1.2). We outline MBH binary merger dynamics in §0.1.3, while a more extensive discussion can be found in §1.3. In §0.2 we describe the basics of GW signal generation by binaries, and detection by pulsar timing arrays. In §0.3 we present the different methods for identifying signatures of MBH binaries electromagnetically. Finally, in §0.4 we briefly outline each thesis chapter.

## 0.1 MASSIVE BLACK HOLES AND GALAXIES

MBH were inferred ([Soltan 1982](#)), and are now observed (e.g. [Kormendy & Richstone 1995](#)), in the majority of massive galaxies. These aren’t chance placements as the mass of the central MBH is strongly correlated with properties of the host galaxy ([Magorrian et al. 1998](#); [Ferrarese & Merritt 2000](#); [Gebhardt et al. 2000](#); [Marconi & Hunt 2003](#))—with MBH typically having a mass  $\sim 10^{-3}$  that of the galaxy’s stellar bulge<sup>†</sup>. Despite the strong correlations, the small relative masses of MBH mean that their gravitational influence on galaxies are weak. Nonetheless, both objects’s dynamics are coupled as the same gaseous inflows which trigger bursts of star formation are also known to trigger MBH growth and AGN activity ([Sanders et al. 1988](#); [Barnes & Hernquist 1992](#)). A model of AGN ‘feedback’, in which emission from the MBH suppresses the further cooling of gas and prevents (or decreases) the efficiency of gas inflow and star formation (e.g. [Di Matteo, Springel & Hernquist 2005](#)), has been gaining traction and has demonstrated marked success in explaining properties of more massive galaxies.

\*‘Hardening’ refers to the process of a binary becoming more tightly bound, in reference to Heggie’s law ([Heggie 1975](#); [Hut 1983](#)).

<sup>†</sup>Which itself is typically a fraction  $\sim 0.1$ – $1.0$  of the total stellar mass (e.g. [Mendel et al. 2014](#)).

The observable emission which defines AGN, often the most luminous in the universe, has long been known as accretion powered (Salpeter 1964; Lynden-Bell 1969). For material to accrete (at some rate:  $dM/dt \equiv \dot{M}$ ) from large scales onto a compact object with radius  $r$ , it must liberate gravitational binding energy at a related rate,

$$\left. \frac{dE}{dt} \right|_{\text{acc}} \equiv L_{\text{acc}} \sim \frac{1}{4} \dot{M} c^2 \left( \frac{r}{R_s} \right)^{-1},$$

where  $R_s \equiv 2GM/c^2$  is the Schwarzschild radius. For black holes, where  $r \sim R_s$ , the energy released is thus a significant fraction of the accretion rest-mass energy\*. We have made no mention of the manner in which this energy is liberated. This agnosticism is often parametrized using  $\varepsilon_{\text{rad}}$ —the radiative efficiency, or the fraction of the rest-mass energy which is effectively radiated away†—i.e.,

$$L_{\text{acc}} = \varepsilon_{\text{rad}} \dot{M} c^2. \quad (1)$$

We can define a characteristic, effective-maximum luminosity—the Eddington luminosity (e.g. Frank, King & Raine 2002)—by equating the radiation force from Thomson scattering (with cross-section  $\sigma_T$ ) to that of gravity,

$$L_{\text{Edd}} \equiv \frac{4\pi GMm_p c}{\varepsilon_{\text{rad}} \sigma_T} \approx 1.3 \times 10^{39} \text{ erg s}^{-1} \left( \frac{M}{M_\odot} \right) \left( \frac{\varepsilon_{\text{rad}}}{0.1} \right)^{-1}. \quad (2)$$

Using (1), the accretion rate powering such a luminosity—the Eddington accretion rate—can be calculated:

$$\dot{M}_{\text{Edd}} \equiv \frac{L_{\text{Edd}}}{c^2} \approx 1.4 \times 10^{19} \text{ g s}^{-1} \left( \frac{M}{M_\odot} \right) \left( \frac{\varepsilon_{\text{rad}}}{0.1} \right)^{-1}. \quad (3)$$

In a seminal paper, Soltan (1982) shows that the total accreted mass required to produce the AGN emission observed over the history of the universe requires a present day mass density comparable to that of the observed MBH population.

More detailed models, often in cosmological simulations which strive to reproduce the statistical properties in large volumes of the universe, have continued to show the accreting-MBH model of AGN well explains not only observational properties of AGN‡ but also many properties of galaxies populations. Ironically, and despite this success, the mechanism of progenitor or ‘seed’ formation for MBH remains unknown (see, e.g., Yu & Tremaine 2002; Ferrarese & Ford 2005). The three primary mechanisms that have been proposed are: 1) rapid accretion onto initially very massive, Population III stars (e.g. Madau & Rees 2001); 2) the direct collapse of pristine gas in protogalaxies (possibly with a phase as a supermassive star; e.g. Bromm & Loeb 2003; Begelman, Volonteri & Rees 2006) or 3) runaway, dynamical collisions between thousands of individual stars in dense stellar systems (e.g. Devecchi & Volonteri 2009). While seed formation continues to be a very active field of study (e.g. Dijkstra, Ferrara & Mesinger 2014), most larger-volume cosmological simulations—like those discussed below—simply assume that MBH seeds form, without specifying a particular mechanism.

### 0.1.1 COSMOLOGICAL SIMULATIONS

By their very nature, the cosmological evolution of the universe is determined by a plethora of phenomena on all time and size scales. Depending on the particular focuses of a given study or simulation,

\*Compare this to typical efficiencies of  $10^{-3}$  and  $10^{-2}$  for fission and fusion respectively.

†The above (Newtonian and virial) estimate suggests an efficiency of  $\sim 1/4$ , while more careful analytic (e.g. Salpeter 1964), observational (e.g. Davis & Laor 2011) and numerical (e.g. Noble et al. 2011; Sądowski et al. 2015) considerations give typical values  $\varepsilon_{\text{rad}} \approx 0.1$  in radiatively efficient accretion flows which we adopt as a fiducial value.

‡Including quasars, liners, blazars, etc; for a review of AGN unification, see, Antonucci (1993).

many different techniques can be used (for a review, see, e.g. [Dolag et al. 2008](#)). On the largest length scales, the universe can be considered as dominated by collisionless cold Dark Matter (DM) in an evolving spacetime. In this approximation, the universe is typically simulated using ‘N-Body’ pure-gravity\* simulations on a background expanding universe (i.e. Friedmann-Lemaitre, [Peebles 1980](#)). Each ‘particle’ in these simulations represents some Monte Carlo realizations of the underlying mass-distribution, instead of a particular, discrete, astrophysical object. Individual close encounters between particles then become largely meaningless, while producing spurious two-body relaxation and requiring short time-steps to accurately resolve. A gravitational smoothing length, typically a small fraction of the average inter-particle spacing, is usually introduced to mediate these effects by damping forces on smaller scales.

Instead of directly calculating the force between each of  $N$  particles, leading to a  $N^2$  computation, either a ‘tree’ ([Barnes & Hut 1986](#)) or ‘particle-mesh’ (PM; [Hockney, Goel & Eastwood 1974](#)) method is typically used. In the tree method, the simulation volume is recursively subdivided (usually into 8 sub-cubes, with side-lengths half that of the parent division), until the lowest ‘leaf’ levels contain only 1 (or some small number) of particles. The force at small distances can be calculated from all nearby particles, while at large distances the force from entire cubes can be used. In general, the tree will be ‘descended’ until the solid angle of a cube is below some critical value determined to balance accuracy and speed. In the PM method, the potential field is discretized onto a grid over which the force can be calculated more efficiently, often in Fourier space. Recent codes typically hybridize between these methods using the ‘TreePM’ approach in which a tree (or even direct N-body) calculation can be used at short-distances, and a PM grid on larger ones (e.g. [Springel 2005](#)).

While DM is the primary mass-constituent of the universe, baryonic matter is more observable. On cosmological scales, most baryons can be treated as a perfect fluid and evolved using conservation equations (mass, momentum and energy; or the Euler, hydrodynamic and thermodynamic equations respectively) along with a suitable equation of state to describe the microscopic behavior of the gas. Two classes of methods are typically distinguished for solving hydrodynamic problems depending on whether the conservation equations are formulated in the Eulerian ‘grid’ methods (considering volume elements) or Lagrangian ‘Smoothed Particles Hydrodynamics’ (SPH) formalism (considering mass elements).

Eulerian methods discretize a fluid onto a grid ([Cen 1992](#); [Ryu et al. 1993](#)), often a regular lattice in the case of a ‘structured’ grid, and use reconstruction techniques (e.g. [Colella & Woodward 1984](#)) to calculate the hydrodynamic quantities as needed. The hydrodynamic equations then become a series of ‘Riemann’ (i.e. discontinuity) problems to calculate fluxes at grid-boundaries ([Courant & Friedrichs 1948](#)). Grid codes are ultimately limited to some finite spatial resolution based on the grid size. They exhibit excellent performance over a tremendous range of densities, however, and can be very effective at capturing discontinuities and shocks which often arise in high-density and super-sonic regions of the universe.

SPH methods are an extension of purely n-body techniques, except particles represent fluid elements ([Gingold & Monaghan 1977](#); [Lucy 1977](#)) instead of purely gravitating masses. One of the strongest advantages of this technique is that for relatively constant-mass elements, high density regions will ‘automatically’ have higher spatial resolutions ([Hernquist & Katz 1989](#)). Fluid parameters are calculated by smoothing over each particle using a kernel function with some characteristic ‘smoothing length’, typically calculated independently for each particle to contain some predetermined number of neighbors (e.g. [Springel & Hernquist 2002](#)). Unfortunately these techniques require an artificial-viscosity term to capture shocks, which also degrades the resolution of discontinuities (e.g. [Monaghan 1992](#)). Additionally, steep density-gradients can produce spurious pressure forces which hinder SPH’s ability to accurately model dynamical instabilities (e.g. [Hernquist 1993](#)).

\*In particular: solving the collisionless Boltzmann equation coupled with the Poisson equation for gravity.

A hybrid approach also exists called a ‘moving-mesh’ which strives to capture the benefits of both SPH and grid-based codes (Fiedler & Mouschovias 1992; Gnedin 1995; Pen 1998). In this technique, the hydrodynamic equations are solved in a fundamentally Eulerian manner: using grids and grid-boundaries which can capture discontinuities and sharp density-gradients. At the same time, the grid boundaries are allowed to deform and align with discontinuities, and cells allowed to move—decrease the velocity gradients across them. The Illustris simulations, which are used extensively throughout this thesis, are based on the *Arepo* moving-mesh code created by Springel (2010). *Arepo* uses a Voronoi tessellation with mesh generating points that are allowed to move with the flow, allowing for full Galilean invariance which drastically improves the accuracy in tracking supersonic elements. A Delaunay triangulation is used to construct cells based on the tessellation, which together allow for quick mesh-construction; smooth, regular cell motions and easy mesh refinement.

### 0.1.2 ILLUSTRIS

The preceding section has discussed the methods of gravitational and purely-hydrodynamic simulations. Galaxies, however, involve a wealth of additional physics, for example, magnetic fields, thermal conduction, radiation transfer & radiative cooling, and star formation (e.g. Dolag et al. 2008). The Illustris simulations (Vogelsberger et al. 2014a,b; Genel et al. 2014; Torrey et al. 2014; Rodriguez-Gomez et al. 2015; Sijacki et al. 2015; Nelson et al. 2015; Snyder et al. 2015) are a suite of simulations in a  $(106.5 \text{ Mpc})^3$  box using *Arepo* to model gas in moving-mesh cells, along with collisionless DM particles, and extensive sub-grid models of cooling, star formation, & feedback (Vogelsberger et al. 2013; Torrey et al. 2014). Three different resolutions were run, each with DM-only and full baryon-physics runs, but we discuss only the highest resolution, full-physics run.

Simulations are initialized at redshift  $z = 127$  with roughly  $6 \times 10^9$  of each collisionless DM particles and hydrodynamic gas cells, where the baryons have masses and softening lengths  $\sim 10^6 M_\odot$  and  $\sim 700 \text{ pc}$  respectively. Gas cooling is calculated as a function of density, temperature, metallicity, the local radiation field of AGN, and the ionizing background radiation (Vogelsberger et al. 2014a). Both primordial and metal line cooling are considered, with corrections for self-shielding. Once gas reaches a critical, star-formation threshold density, it is given an effective equation-of-state representing a two-phase ISM of cool clouds in pressure equilibrium with a hot component (Springel & Hernquist 2003). Star formation is implemented stochastically following a local Kennicutt-Schmidt law and using a Chabrier IMF to construct a coeval population represented by collisionless star particles (Vogelsberger et al. 2014b). Each star particle tracks the evolution of stars through AGB outflows and type Ia & type II supernovae. Outflows and supernovae blow metal-enriched gas—with nine different species being tracked—into the surrounding medium. Stellar feedback is implemented by coupling thermal energy to the local gas which prevents a cooling runaway and self-regulates the star formation rate. The full evolution of stars and galaxy over time is presented in detail in Genel et al. (2014).

Here we summarize some of the important, massive black-hole models in Illustris, but we refer the reader to Di Matteo, Springel & Hernquist (2005) and Sijacki et al. (2007) for details on the prescriptions, and to Sijacki et al. (2015) for the populations and properties of the MBH resulting from Illustris. A friends-of-friends algorithm (Davis et al. 1985) is used to identify DM halos as the simulation progresses. MBH are seeded with a mass of  $\sim 10^5 M_\odot$  into halos once they reach  $7 \times 10^{10} M_\odot$ . To prevent spurious wandering of MBH particles due to scattering events, the MBH particles are manually repositioned to their potential minima of their galaxies, which, unfortunately invalidates their velocity and detailed-

positional information. MBH particles grow both through Bondi-like accretion\* from the nearby gas, and by MBH ‘mergers’ which occur instantly once two MBH particles come within a smoothing length of one-another ( $\sim$  kpc). As MBH accrete, they feedback on the environment in three ways. Quasar-mode feedback, which occurs for Eddington ratios  $f_{\text{Edd}} \geq 0.05$ , uses a bolometric luminosity calculated from a fixed radiative efficiency ( $\varepsilon_{\text{rad}} = 0.2$ ) to deposit a fraction of energy (5%) into the surrounding gas. Radio-mode feedback, active for  $f_{\text{Edd}} < 0.05$ , deposits a fraction (35%) of the luminosity in randomly placed ‘bubbles’ surrounding the MBH. Finally, radiative-feedback alters the ionization state of local gas which directly changes its cooling rate.

At the end of the simulation, the volume contains some  $6 \times 10^9$  star particles, and 32,000 MBH particles with a history of almost 24,000 merger events †. The match between Illustris and observations are striking (Vogelsberger et al. 2014a). Galaxy stellar-mass and stellar-luminosity functions along with star-forming and quenched galaxy-fractions are in excellent agreement with SDSS galaxies, and star formation efficiencies are consistent with abundance matching measurements. The MBH population shows good agreement with mass-density and mass-function estimates, and derived bolometric and hard x-ray luminosity functions match AGN observations (Sijacki et al. 2015). Post-processing catalogs of galaxy merger-rates (Rodriguez-Gomez et al. 2015) in addition to galaxy spectra and photometry (Torrey et al. 2014) have also been produced. These, along with the raw simulation data, are openly available online (www.illustris-project.org; Nelson et al. 2015).

### 0.1.3 BINARY DYNAMICS

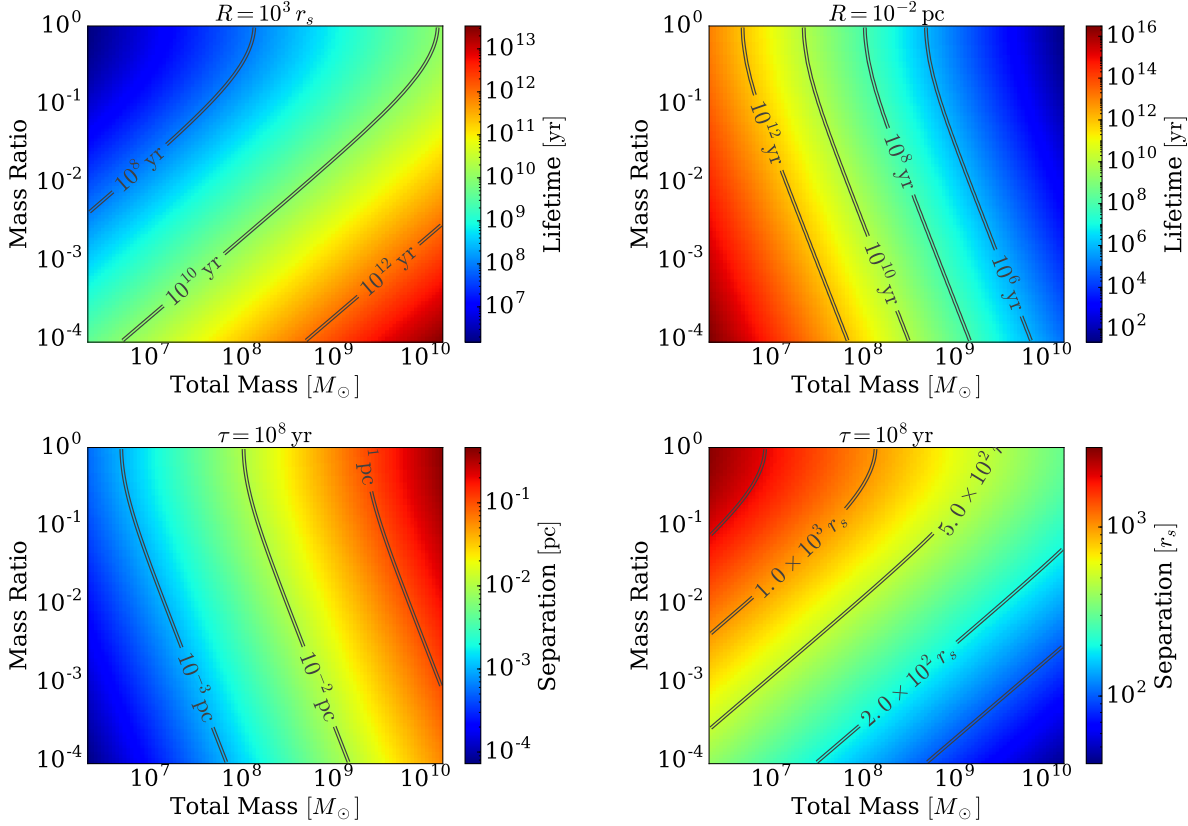
Mergers between galaxies are a fundamental aspect of hierarchical structure formation (Blumenthal et al. 1984; White & Frenk 1991), and have been extensively observed (Zwicky 1956; Lotz et al. 2011) and seen in simulations (Lacey & Cole 1993; Rodriguez-Gomez et al. 2015). This presents two possibilities for the pair of MBH finding themselves in a merger-remnant: either they also merge, or they persist as multiples. Because Illustris MBH are manually ‘coalesced’ at  $\sim$  kpc scales, the dynamics of the true merger process are almost entirely unresolved. There is strong, theoretical motivation for MBH to merge effectively (see below and §1.3, e.g. Hayasaki 2009; Preto et al. 2011; Khan et al. 2013; Goicovic et al. 2017). Additionally, one might naively expect that MBH *must* undergo mergers to grow, but this isn’t *necessarily* the case. In particular, the linear growth provided by mergers is known to be woefully inadequate to form the massive quasars observed at high redshifts (e.g. Fan et al. 2006), which require exponential growth from accretion (e.g. Haiman 2013) to fully blossom. Together with the aforementioned connection between the accretion history of AGN, and the present day mass-content of MBH (Soltan 1982), it would seem that binary coalescences are neither necessary nor sufficient to explain most MBH observations (Small & Blandford 1992).

The fate and physics of MBH following a galaxy-merger was first outlined in a pioneering study by Begelman, Blandford & Rees (1980). The authors describe how the merger progresses through four key stages of evolution. **(1)** On large scales ( $\sim$  kpc) MBH (and indeed the two merging galaxies as a whole) are brought together predominantly by dynamical friction, until they meet in the galactic center and form a binary. **(2)** At some point, individual three-body scatterings between stars and the binary must be considered, including depletion of the stars with sufficiently low angular-momentum to interact with the bi-

\*We specify Bondi-like, because while  $\dot{M} \propto \rho M^2/c_s^3$ , where the local density and sound speed are  $\rho$  and  $c_s$ , as in Bondi (Bondi 1952), the normalization is rescaled to match observed MBH-galaxy scaling relations.

†Though many of these occur for MBH near their seed-mass and may not be entirely physical, see §1.2.1. In most of our analyses we use rigorous selection cuts to identify both reliable mergers in well-resolved systems. This usually leads to a population of  $\sim 10,000$  mergers.

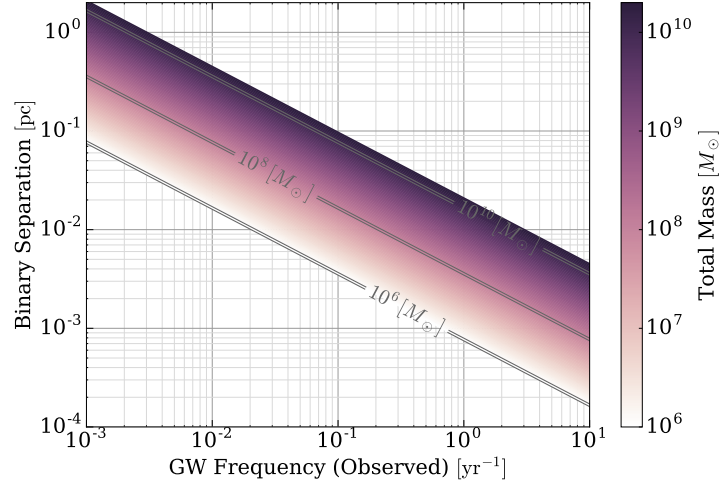




**Figure 1:** Characteristic lifetimes and separations for coalescence due to purely GW emission over the relevant parameter space of mass ratio and total mass. The upper panels are colored by time to coalescence for fixed separations:  $R = 10^3 R_s$  (left), and  $R = 10^{-2} \text{ pc}$  (right). The lower panels are colored by separation to coalesce within  $10^8 \text{ yr}$ , in units of parsecs (left) and Schwarzschild radii ( $R_s$ , right).

nary: the so-called, ‘Loss Cone’ (LC; for a review, see, Merritt & Milosavljević 2005). Using characteristic quantities, if the LC is effectively replenished (i.e. remains ‘full’), binaries are generally able to continue merging. If, on the other hand, the LC subsists at a depleted level, binaries will stall: the so-called ‘final-parsec’ problem (see: Merritt 2013). **(3)** For binaries in gas-rich systems, which reach sufficiently small scales ( $\lesssim 0.1 \text{ pc}$ ), gas-drag can contribute significantly to the hardening process. **(4)** Eventually GW emission inevitably dominates at the smallest scales ( $\lesssim 10^{-3} \text{ pc}$ ), finally driving binaries to coalescence. Each of these processes are discussed in detail in §1.3.

To get a sense of the scales at which gravitational wave emission becomes important, Fig. 1 shows the GW-inspiral timescales at fixed separation (top row), and the separations to coalesce within a fixed time (bottom row), for the relevant parameter space of total mass and mass ratio. The left column gives separations in terms of Schwarzschild radii ( $R_s = R_s(M = M_1 + M_2)$ ), and the right column in parsec. A typical binary, for example  $M \sim 10^7 M_\odot$  and  $q \sim 0.1$ , must reach  $\sim 2 \times 10^{-3} \text{ pc} \sim 3 \times 10^3 R_s$  to merger within  $10^9 \text{ yr}$ , while a binary more characteristic of a prospective GW source, for example  $M \sim 10^9$  and  $q \sim 0.3$ , need only reach  $\sim 0.1 \text{ pc} \sim 10^3 R_s$ .



**Figure 2:** Binary separation versus gravitational wave frequency for the total MBHB masses used in our simulations. The GW frequency for circular orbits is twice the orbital frequency. For the PTA band, roughly  $0.1\text{--}1\text{ yr}^{-1}$ , binaries are observable between about  $10^{-3}$  and  $10^{-1}$  pc. Most of the GWB amplitude come from higher mass systems, towards larger separations.

## 0.2 GRAVITATIONAL WAVE SIGNALS & PULSAR TIMING ARRAYS

Binaries in circular orbits emit gravitational waves at twice their orbital frequency, with a strain amplitude given by (e.g. [Sesana, Vecchio & Colacino 2008](#)),

$$h_{s,\text{circ}}(f_r) = \frac{8}{10^{1/2}} \frac{(GM)^{5/3}}{c^4 d_L} (2\pi f_r)^{2/3} \sim 10^{-15} \left( \frac{M}{10^9 M_\odot} \right)^{5/3} \left( \frac{d_L}{1 \text{ Gpc}} \right)^{-1} \left( \frac{f_r}{1 \text{ yr}^{-1}} \right), \quad (4)$$

for a luminosity distance  $d_L$ , and rest-frame orbital frequency  $f_r$ . The amplitude of GW from a binary is strongly effected by the ‘chirp-mass’,

$$\mathcal{M} = \frac{(M_1 M_2)^{3/5}}{M^{1/5}} = M \frac{q^{3/5}}{(1+q)^{6/5}}, \quad (5)$$

where the primary and secondary masses are  $M_1$  and  $M_2 < M_1$ , the total mass  $M \equiv M_1 + M_2$ , and the mass-ratio  $q \equiv M_2/M_1$ . To get a sense of typical scales, the relation between orbital period and separation ([Kepler 1619](#)), is shown in Fig. 2 for a variety of relevant total masses. When gravitational-wave emission is the dominant hardening mechanism, the time a binary spends at a given separation—called the ‘hardening time’ or ‘residence time’—becomes a steep power of the frequency (e.g. [Enoki & Nagashima 2007](#)),  $\tau_{\text{gw}} \propto f_r^{8/3}$ . This leads to a rapid acceleration in inspiral which produces the characteristic ‘chirp’ in frequency as a binary coalesces. For typical MBH binaries, the chirp occurs at frequencies of  $\sim 10^{-6}$  Hz corresponding to periods of  $\sim$  days. For populations of systems, the strong frequency-dependence also means that the number density of systems in a given volume of the universe drops rapidly as a function of frequency.

GW signals from astrophysical binaries are typically grouped into three categories: **(1)** ‘deterministic’/‘continuous’ sources—individual binary systems which tend to emit monochromatically\*; **(2)** the stochastic gravitational-wave background (GWB)—the superposition of many binaries all emitting at comparable amplitudes and frequencies; and **(3)** ‘bursts’, where a single peak of emission (or small number of cycles) is observed. A particular subclass of burst sources, so-called ‘bursts with memory’ ([Braginskii](#)

\*The residence time of the MBHB systems we expect to be detectable as deterministic sources are typically thousands to millions of years, meaning that they evolve very little in frequency over typical observing spans.

& Thorne 1987), are especially relevant to PTA (see, e.g., Pshirkov, Baskaran & Postnov 2010; Cordes & Jenet 2012; Arzoumanian et al. 2015), but in this thesis we focus only on the former two classes of signal. As mentioned above, the number of binaries in the universe tends to increase at lower frequencies (where the hardening time is longer). This leads to the GWB becoming dominant at lower frequencies, and deterministic sources appearing predominantly at higher frequencies. In a seminal paper, Phinney (2001) showed that the characteristic GWB spectrum can be calculated analytically by considering a cosmological population of binaries evolving purely due to GW emission, and integrating the total energy emitted as gravitational waves over the history of the universe. This, elegant calculation, which is identical in methodology to that of Soltan (1982), yields a simple spectrum of the form,

$$h_c(f) = A_0 \left( \frac{f}{f_0} \right)^{-2/3}. \quad (6)$$

The power-law relation in (6) has become typical for reporting GWB predictions and upper-limits, and is usually normalized to  $f_0 = 1 \text{ yr}^{-1}$ , with an amplitude  $A_{\text{yr}^{-1}}$ . Predictions for the GWB are typically calculated using semi-analytic models (e.g. Wyithe & Loeb 2003; Jaffe & Backer 2003), though Monte Carlo realizations of hierarchical cosmologies have become popular (Sesana et al. 2004). Predicted amplitudes almost always lie between  $\log A_{\text{yr}^{-1}} = -16$  and  $\log A_{\text{yr}^{-1}} = -14.3$ , with models above  $\log A_{\text{yr}^{-1}} \gtrsim -15$  strongly disfavored by the most recent upper-limits from pulsar timing arrays (e.g. Shannon et al. 2015). An extensive list of GWB amplitude predictions are included in §1.1.

PTA are sensitive to Gravitational Waves from MBH binaries at nanohertz frequencies (Sazhin 1978; Detweiler 1979; Romani & Taylor 1983; Foster & Backer 1990). Three independent PTA are currently in operation: the European (EPTA, Kramer & Champion 2013; Desvignes et al. 2016), NANOGrav (McLaughlin 2013; The NANOGrav Collaboration et al. 2015), and Parkes (PPTA, Manchester et al. 2013; Reardon et al. 2016). Additionally, the International PTA (IPTA, Hobbs et al. 2010; Verbiest et al. 2016) is a collaboration between all three which uses their combined data to boost sensitivity and better characterize noise and other systematics. The characteristic sensitivity of a PTA can be estimated as (Rajagopal & Romani 1995),

$$h^{\text{PTA}} \sim \frac{\sigma_N}{P_{\text{GW}}} \left( \frac{T_{\text{obs}}}{\Delta t_{\text{obs}}} \right)^{-1/2}, \quad (7)$$

for a pulsar-timing RMS  $\sigma_N$ , GW period  $P_{\text{GW}}$ , totally observing during  $T_{\text{obs}}$ , and observing cadence  $\Delta t_{\text{obs}}$ . This shows that a PTA's most-sensitive frequency is its lowest,  $f_{\text{min}}$ : determined by Nyquist sampling from the total observation duration,  $P_{\text{GW,max}} = 1/f_{\text{min}} \sim T_{\text{obs}}$ , and thus,

$$h^{\text{PTA}}(f_{\text{min}}) \sim \sigma_N \Delta t_{\text{obs}}^{1/2} T_{\text{obs}}^{-3/2} \approx 10^{-15} \left( \frac{\sigma_N}{10 \text{ ns}} \right) \left( \frac{\Delta t_{\text{obs}}}{0.05 \text{ yr}} \right)^{1/2} \left( \frac{1 \text{ yr}}{T_{\text{obs}}} \right)^{-3/2}. \quad (8)$$

This scaling is most applicable to the upper limits that can be made on GW strain by a single pulsar. To truly make a detection, the cross-correlations of pulse time-of-arrival variations from many pulsars must be shown to properly correlate across the sky, according to a pattern often referred to as the ‘Hellings & Downs curve’ following Hellings & Downs (1983). The induced timing residual from a particular source is given by the relation (e.g. Sesana, Vecchio & Volonteri 2009),

$$\delta t = \frac{h_s}{f} (T_{\text{obs}} f)^{1/2} = \frac{h_c}{f}. \quad (9)$$

Equation (9) can be taken as the definition of the ‘characteristic strain’  $h_c$ , which takes into account the number of cycles over which a source is observed.

### 0.3 ELECTROMAGNETIC OBSERVATIONS

Galaxy mergers have been observed and identified since the beginnings of extragalactic astronomy, and systems in which both galaxies have an observable AGN are not uncommon (e.g. Zwicky 1956). While inferring or directly observing the presence of an MBH Binary in a galaxy is difficult, many techniques have been developed (see, e.g., Komossa 2006). These methods can be categorized into four groups: spatial offsets, kinematic offsets, morphological indicators, periodic variability.

Spatial offset detection relies on having the angular resolution to discern a pair of ‘dual AGN’ (e.g. Komossa et al. 2003; Comerford et al. 2009b; Liu et al. 2014) or a single ‘offset AGN’ relative to its host galaxy or stellar core (e.g. Comerford & Greene 2014). We use the term ‘dual’ to refer to an associated pair, distinct from a gravitationally bound ‘binary’. Overall,  $\sim 10^{-3}$  of SDSS quasars could be dual AGN (Foreman, Volonteri & Dotti 2009). Large enough samples are beginning to be discovered to begin placing constraints on population statistics, for example showing that dual-AGN occurrence rates are higher in more-major mergers and luminosities tend to increase at smaller separations (Koss et al. 2012)—both consistent with predictions for merger-driven fueling of AGN (e.g. Sanders et al. 1988; Barnes & Hernquist 1992). The current record holder for closest-separation system is by Rodriguez et al. (2006) in the radio galaxy 0402+379, with a projected separation of  $\sim 7$  pc. Even in this case, it is unclear if the system is dynamically bound.

Kinematic offsets are identified spectroscopically, and again can be either ‘dual’ or ‘offset’. Kinematically, a ‘dual’ system has observable lines from two components (double-lines; Comerford et al. 2009a; Eracleous et al. 2012; Tsalmantza et al. 2011). ‘Offset’ kinematic systems have only single lines observable, but at a noticeable redshift (i.e. velocity offset) relative to the host or other component (e.g. Gaskell 1983). While Wang et al. (2009), for example, reports that as many as  $10^{-2}$  of SDSS AGN could be dual, occurrence rates must be carefully vetted as the kinematic technique has seen numerous examples of false-positives (e.g. Halpern & Filippenko 1988; Eracleous et al. 1997). Unfortunately, double lines can also be produced by bipolar outflows or jets and even the two sides of an accretion disk (Shen et al. 2011). Many recent studies have adopted an approach of identifying candidates through double-peaked emission lines (particularly, [OIII]), and then following them up in search of additional signatures, for example tidal features and multiple stellar components (Liu et al. 2010; Fu et al. 2011a), or spatially offset AGN (Fu et al. 2011b; Barrows et al. 2016). These multi-method identification surveys will hopefully provide insights into a growing number of small-separation candidates identified via kinematic offsets (e.g. Boroson & Lauer 2009; Dotti et al. 2009).

Morphological indicators come in numerous forms. Helical structure in radio jets has been used to suggest the ongoing presence of a binary (e.g. Roos, Kaastra & Hummel 1993). ‘X-Type’ radio sources (e.g. Ekers et al. 1978), on the other hand, have been explained as the rapid reorientation of an AGN jet following its coalescence with another MBH (Merritt & Ekers 2002). Another possible indicator of past binarity is a decrease in the surface brightness of some galaxies towards their core. A decreased core stellar-density may be due to evacuation or ‘coring’ from a binary inspiral (Lauer et al. 2002), as large numbers of stars can be ejected by the system as they extract orbital energy from the binary. Morphological indicators, perhaps even more so than kinematic offsets, can be somewhat ambiguous by themselves, but can certainly bolster the binary hypothesis in systems with additional multiplicity indicators.

Periodic variability surveys examine the brightness of AGN in search of regular modulations which could indicate the presence of a companion\* (Fan et al. 2002; De Paolis, Inghrosso & Nucita 2002; Xie et al.

\*Periodic variations of velocity-offset features are also important, but we consider those an extension of the kinematic method.

2002). One particular system, OJ287, has shown remarkably consistent outbursts every 12 yr, over a 100 yr baseline, which has been taken as very strong evidence for an eccentric MBHB system, in which the secondary periodically punctures the disk of the primary (Valtonen et al. 2008). PG 1302-102 (Graham et al. 2015b), is another provocative example of roughly two cycles of sinusoidal variations, in this case, which can be well explained by doppler-boosting from a mildly relativistic orbital velocity (D’Orazio, Haiman & Schiminovich 2015). Recent followup however, which extends the observing baseline, fails to see the periodicity continue Liu, Gezari & Miller 2018. This highlights the dangers of misidentifying red-noise processes, which appear sinusoidal over poorly-sampled periods, as a truly periodic physical process. One complicating factor is that while red-noise is expected to become more prevalent at longer periods, so is the overall occurrence rate of binaries—again due to the strong frequency dependence in hardening rate.

Recently, numerous large-scale surveys have systematically searched for periodic variables in their AGN samples: Graham et al. (2015a) find 111 candidates in  $\sim 240,000$  AGN using the CRTS survey; Charisi et al. (2016) find 33 in  $\sim 35,000$  AGN using PTF; and Liu et al. (2016) initially identify 3 candidates in 670 AGN using PanSTARRS, however, none are persistent after adding archival data. If the variability timescales of  $\sim$ yr in the CRTS and PTF candidates are associated with the orbital period of an MBHB, it means their GW signals would be in the PTA sensitive band—making these an especially exciting population.

## 0.4 THESIS PREVIEW

This thesis aims to make comprehensive predictions for the populations, lifetimes, and environments of MBH binaries. Using these data, we calculate the predicted properties of both the GW background and foreground (deterministic sources), along with expected times to detection using realistic models of pulsar timing arrays. We additionally explore the consistency between recently identified periodically-variable, MBHB candidates and the current upper limits on GW signals from PTA. We then explore predictions for the occurrence rates and parameter distributions of periodic-variable systems based on different models for producing their variability.

In Ch. 1 (Kelley, Blecha & Hernquist 2017), we present our population of MBH binaries and their host galaxies, in addition to their evolution based on comprehensive semi-analytic models of the merger process. We give merger lifetimes and discuss the properties of systems able to effectively merge within a Hubble time, in addition to basic predictions of the GWB amplitude based on our data. Chapter 2 (Kelley et al. 2017) discusses the addition of eccentric binary-evolution to our models, and is primarily dedicated to a careful calculation of the GWB using a much more accurate, Monte-Carlo method. In addition to exploring how eccentric and environmental effects impact the GWB spectrum, we use models of the four existing PTA to calculate expected times-to-detection—finding very promising results. In Ch. 3 (Kelley et al. 2018a), we analyze our MBHB as individually-resolvable, deterministic GW sources. We present the statistical properties of typical GW strain amplitudes and pulsar timing-residuals, as a function of evolutionary parameters. In this section we also carefully examine times-to-detection, and find that our results are different from previous studies. We also present the first careful analysis of varying pulsar red-noise properties on detection prospects—which remain a highly uncertain aspect of PTA observations.

Electromagnetic observations are the focus of Ch. 4 (Kelley et al. 2018b), in which we predict the detection rates and parameter distributions of periodic-variable AGN, based on our simulated MBHB. Along with the results of Sesana et al. (2017a), we use these models to compare between the large number of candidate MBHB recently identified in numerous optical surveys, with the current most stringent upper limits on MBHB GW signatures by PTA. In our conclusions (§5), we summarize our most important results and

their impact on future work. We close by discussing numerous followup studies which are of particular interest.

# MASSIVE BLACK HOLE BINARY MERGERS IN DYNAMICAL GALACTIC ENVIRONMENTS

*This thesis chapter originally appeared in the literature as*

Luke Zoltan Kelley, Laura Blecha, Lars Hernquist  
MNRAS, 464, 3. arXiv:1606.01900

## ABSTRACT

GRAVITATIONAL Waves (GW) have now been detected from stellar-mass black hole binaries, and the first observations of GW from Massive Black Hole (MBH) Binaries are expected within the next decade. Pulsar Timing Arrays (PTA), which can measure the years long periods of GW from MBHB, have excluded many standard predictions for the amplitude of a stochastic GW Background (GWB). We use coevolved populations of MBH and galaxies from hydrodynamic, cosmological simulations ('Illustris') to calculate a predicted GWB. The most advanced predictions so far have included binary hardening mechanisms from individual environmental processes. We present the first calculation including all of the environmental mechanisms expected to be involved: dynamical friction, stellar 'loss-cone' scattering, and viscous drag from a circumbinary disk. We find that MBH binary lifetimes are generally multiple gigayears, and only a fraction coalesce by redshift zero. For a variety of parameters, we find all GWB amplitudes to be below the most stringent PTA upper limit of  $A_{\text{yr}^{-1}} \approx 10^{-15}$ . Our fairly conservative fiducial model predicts an amplitude of  $A_{\text{yr}^{-1}} \approx 0.4 \times 10^{-15}$ . At lower frequencies, we find  $A_{0.1\text{yr}^{-1}} \approx 1.5 \times 10^{-15}$  with spectral indices between  $-0.4$  and  $-0.6$ —significantly flatter than the canonical value of  $-2/3$  due to purely GW-driven evolution. Typical MBHB driving the GWB signal come from redshifts around 0.3, with total masses of a few times  $10^9 M_{\odot}$ , and in host galaxies with very large stellar masses. Even without GWB detections, our results can be connected to observations of dual AGN to constrain binary evolution.

## 1.1 INTRODUCTION

Massive Black Holes (MBH) occupy at least the majority of massive galaxies (e.g. Soltan 1982; Kormendy & Richstone 1995; Magorrian et al. 1998) which are also known to merge with each other as part of their typical lifecycles (e.g. Lacey & Cole 1993; Lotz et al. 2011; Rodriguez-Gomez et al. 2015). This presents two possibilities for the MBH of host galaxies which merge: either they also merge, or they persist as multiples in the resulting remnant galaxies. Naively, one might expect that BH *must* undergo merg-

ers for them to grow—as for halos and to some degree galaxies in the fundamentally hierarchical Lambda-Cold Dark Matter model (e.g. [White & Frenk 1991](#)). On the contrary, the linear growth of black holes (i.e. at most doublings of mass) is known to be woefully inadequate to form the massive quasars observed at high redshifts (e.g. [Fan et al. 2006](#)), which require exponential growth from Eddington\* (or super-Eddington) accretion (e.g. [Haiman 2013](#)). Assuming that the energy fueling Active Galactic Nuclei (AGN) comes from accretion onto compact objects, the integrated luminosity over redshifts requires a present day mass density comparable to that of observed MBH ([Soltan 1982](#)). Coalescences of MBH are then neither sufficient nor necessary to match their observed properties (e.g. [Small & Blandford 1992](#)).

In the last decade the search for multi-MBH systems have yielded many with kiloparsec-scale separations (‘Dual AGN’, e.g. [Comerford et al. 2012](#))—although they seem to represent only a fraction of the population ([Koss et al. 2012](#))—and even some triple systems (e.g. [Deane et al. 2014](#)). At separations of kiloparsecs, ‘Dual MBH’ are far from the ‘hard’ binary phase. A ‘hard’ binary is one in which the binding energy is larger than the typical kinetic energy of nearby stars ([Binney & Tremaine 1987](#)), and relatedly the binding energy tends to increase (i.e. the system ‘hardens’) with stellar interactions ([Hut 1983](#)). ‘Hard’ is also used more informally to highlight systems which behave dynamically as a bound system.. Only one true MBH Binary (MBHB, i.e. gravitationally bound) system has been confirmed: a resolved system in the radio galaxy 0402+379, at a separation of  $\sim 7$  pc by [Rodriguez et al. \(2006\)](#). Even this system is well outside of the ‘Gravitational Wave (GW) regime’, in which the system could merge within a Hubble time due purely to GW emission. There are, however, a growing number of candidate, unresolved systems with possible sub-parsec separations (e.g. [Valtonen et al. 2008](#); [Dotti et al. 2009](#)). Detecting—and even more so, excluding—the presence of MBHB is extremely difficult because activity fractions of AGN (particularly at low masses) are uncertain, the spheres of influence of MBH are almost never resolved, and the expected timescales at each separation are unknown. It is then hard to establish if the absence of MBHB observations is conspicuous.

On the theoretical side, the picture is no more complete. Pioneering work by [Begelman, Blandford & Rees \(1980, hereafter BBR80\)](#) outlined the basic MBH merger process. On large scales ( $\sim$  kpc), MBH are brought together predominantly by dynamical friction—the deceleration of a body moving against a gravitating background. Energy is transferred from the motion of the massive object to a kinetic thermalization of the background medium, in this case the dark-matter, stellar and gaseous environment of MBHB host galaxies. As the binary tightens, stars become the primary scatterers. Once the binary becomes hard ( $\sim 10$  pc), depletion of the ‘Loss Cone’ (LC)—the region of parameter space with sufficiently low angular momentum to interact with the MBHB—must be considered. The rate at which the LC is refilled is likely the largest uncertainty in the merger process, and determines the fraction of systems which are able to cross the so-called ‘final-parsec’ (for the definitive review, see: [Merritt 2013](#)). If binaries are able to reach smaller scales ( $\lesssim 0.1$  pc), gas-drag can contribute significantly to the hardening process in gas-rich systems. Eventually GW emission inevitably dominates at the smallest scales ( $\lesssim 10^{-3}$  pc).

Since [BBR80](#), the details of the merger process have been studied extensively, largely focusing on the LC (for a review, see, e.g. [Merritt & Milosavljević 2005](#)). [Quinlan \(1996\)](#) and [Quinlan & Hernquist \(1997\)](#) made significant developments in numerical, N-Body scattering experiments, allowing the ‘measurement’ of binary hardening parameters. More advanced descriptions of the LC, applied to realistic galaxy density profiles, by [Yu \(2002\)](#) highlighted the role of the galactic gravitational potential—suggesting that flattened

\*The Eddington accretion rate can be related to the Eddington luminosity as  $\dot{M}_{\text{Edd}} = L_{\text{Edd}}/\epsilon_{\text{rad}}c^2$ , for a radiative efficiency  $\epsilon_{\text{rad}}$ , i.e.

$$\dot{M}_{\text{Edd}} = \frac{4\pi GMm_{\text{p}}}{\epsilon_{\text{rad}}c\sigma_{\text{T}}},$$

for a proton mass  $m_{\text{p}}$ , and Thomson-Scattering cross-section  $\sigma_{\text{T}}$ .



and strongly triaxial galaxies could be effective at refilling the LC and preventing binaries from stalling. N-Body simulations have continued to develop and improve our understanding of rotating galaxies (e.g. [Berczik et al. 2006](#)), and more realistic galaxy-merger environments and galaxy shapes (e.g. [Khan, Just & Merritt 2011](#); [Khan et al. 2013](#)). The interpretation and usage of simulations have also developed significantly, in step with numerical advancements, allowing for better understanding of the underlying physics (e.g. [Sesana & Khan 2015](#); [Vasiliev, Antonini & Merritt 2015](#)). Smoothed Particle Hydrodynamic (SPH) simulations of MBH binary dynamics in gaseous environments have also been performed. [Escala et al. \(2005a,b\)](#) showed that dense gaseous regions, corresponding to ULIRG-like\* galaxies, can be very effective at hardening binaries. Similar SPH studies have affirmed and extended these results to more general environments and MBHB configurations (e.g. [Dotti et al. 2007](#); [Cuadra et al. 2009a](#)). While modern simulations continue to provide invaluable insights, and exciting steps towards simulating MBHB over broad physical scales are underway (e.g. [Khan et al. 2016](#)), neither hydrodynamic nor purely N-Body simulations are close to simulating the entire merger process in its full complexity<sup>†</sup>.

With our entrance into the era of Gravitational Wave (GW) astronomy ([The LIGO Scientific Collaboration & the Virgo Collaboration 2016](#)), we are presented with the prospect of observing compact objects outside of both the electromagnetic spectrum and numerical simulations. Direct detection experiments for gravitational waves are based on precisely measuring deviations in path length (via light-travel times). While ground-based detectors like the Laser Interferometer Gravitational-Wave Observatory (LIGO; [LIGO 2016](#)) use the interference of light between two orthogonal, kilometer scale laser arms, Pulsar Timing Arrays (PTA, [Foster & Backer 1990](#)) use the kiloparsec scale separations between earth and galactic pulsars ([Detweiler 1979](#)). PTA are sensitive to GW at periods between the total observational baseline and the cadence between observations. These frequencies, roughly  $0.1 - 10 \text{ yr}^{-1}$ , are much lower than LIGO—corresponding to steady orbits of MBHB with total masses between  $\approx 10^6 - 10^{10} M_{\odot}$ , at separations of  $\approx 10^{-3} - 10^{-1} \text{ pc}$  (i.e.  $1 - 10^6 R_s$ <sup>‡</sup>). The parameter space is shown in [Fig. 2](#).

Binaries produce GW which increase in amplitude and frequency as the orbit hardens, up to the ‘chirp’ when the binary coalesces. MBHB chirps will be at frequencies below the LIGO band, but above that of PTA. Future space-based interferometers (e.g. eLISA; [The eLISA Consortium et al. 2013](#)) will bridge the divide and observe not only the coalescence of MBHB, but also years of their final inspiral. The event rate of nearby, hard MBHB which could be observed as individual ‘continuous wave’ sources is expected to be quite low, and likely the first GW detections from PTA will be of a stochastic GW Background (GWB) of unresolved sources ([Rosado, Sesana & Gair 2015](#)).

The shape of the GWB spectrum was calculated numerically more than two decades ago ([Rajagopal & Romani 1995](#)), but [Phinney \(2001\)](#) showed that the characteristic GWB spectrum can be calculated analytically by considering the total energy emitted as gravitational waves, integrated over redshift. For a complete and pedagogical derivation of the GWB spectrum see, e.g., [Sesana, Vecchio & Colacino \(2008\)](#). The ‘characteristic-strain’,  $h_c(f)$ , can be calculated for a finite number of sources, in some comoving volume  $V_c$  (e.g. a computational box), as,

$$h_c^2(f) = \frac{4\pi}{3c^2} (2\pi f)^{-4/3} \sum_i \frac{1}{(1+z_i)^{1/3}} \frac{(GM_i)^{5/3}}{V_c}. \quad (1.1)$$

Equation 1.1 is the simplest way to calculate a GW background spectrum, requiring just a distribution of

\*Ultra-Luminous Infra-Red Galaxies (ULIRG) are bright, massive, and gas-rich — all indicators of favorable MBH merger environments.

<sup>†</sup>Resolving the interaction of individual stars with a MBHB over the course of the entire merger process, for example, would require almost nine orders of magnitude contrast in each mass, distance, and time.

<sup>‡</sup>Schwarzschild radii,  $R_s \equiv 2GM/c^2$ .

merger chirp-masses and redshifts. This type of relation is often written as,

$$h_c(f) = A_0 \left( \frac{f}{f_0} \right)^{-2/3}, \quad (1.2)$$

which has become typical for GWB predictions, and usually normalized to  $f_0 = 1 \text{ yr}^{-1}$  (with some  $A_{\text{yr}^{-1}}$ ). The prediction of a GWB with spectral slope of  $-2/3$  is quite general, but does assume purely gravitational-wave driven hardening which produces a purely power-law evolution in frequency. The lack of high and low frequency cutoffs is fortuitously accurate at the frequencies observable through PTA, which are well populated by astrophysical MBHB systems. Deviations from pure power-law behavior within this band, however, are not only possible but expected—the degree of which, determined by how significant non-GW effects are, is currently of great interest.

Many predictions have been made for the normalization of the GWB based on extensions to the method of [Phinney \(2001\)](#). The standard methodology is using Semi-Analytic Models\* (SAM) of galaxy evolution, with prescribed MBHB mergers to calculate a GWB amplitude. Two of the earliest examples are [Wyithe & Loeb \(2003\)](#)—who use analytic mass functions ([Press & Schechter 1974](#)) with observed merger rates ([Lacey & Cole 1993](#)), and [Jaffe & Backer \(2003\)](#)—who use observationally derived galaxy mass functions, pair fractions, and merger time-scales. These studies find amplitudes of  $\log A_{\text{yr}^{-1}} = -14.3$  and  $\log A_{\text{yr}^{-1}} = -16$ , respectively which remain as upper and lower bounds to most predictions since then. Monte Carlo realizations of hierarchical cosmologies ([Sesana et al. 2004](#)) exploring varieties of MBHB formation channels (e.g. [Sesana, Vecchio & Colacino 2008](#); [Sesana 2013b](#); [Roebber et al. 2016](#)) have been extremely fruitful in populating and understanding the parameter space, finding GWB amplitudes generally consistent with  $\log A_{\text{yr}^{-1}} \approx -15 \pm 1$ . [Sesana et al. \(2016\)](#) find that accounting for bias in MBH-Host scaling relations moves SAM predictions towards the lower end of this range at  $\log A_{\text{yr}^{-1}} = -15.4$ .

More extensive models exploring deviations from the purely power-law GWB have also been explored. For example, at higher frequencies ( $\gtrsim 1 \text{ yr}^{-1}$ ) from a finite numbers of sources ([Sesana, Vecchio & Volonteri 2009](#)), or at lower frequencies due to eccentric binary evolution (e.g. [Sesana 2010](#)). Recently, much work has focused on the ‘environmental effects’ outlined by [BBR80](#). [Kocsis & Sesana \(2011\)](#) incorporate viscous drag from a circumbinary gaseous disk ([Haiman, Kocsis & Menou 2009](#), hereafter [HKM09](#)) on top of halos and mergers from the dark-matter only Millennium simulations ([Springel et al. 2005](#)), with MBH added in post-processing. They find a fairly low amplitude GWB,  $\log A_{\text{yr}^{-1}} \approx -16 \pm 0.5$ , with a flattening spectrum below  $\sim 1 \text{ yr}^{-1}$ . [Ravi et al. \(2014\)](#) explore eccentric binary evolution in an always effectively refilled (i.e. full) LC using the Millennium simulation with the SAM of [Guo et al. \(2011\)](#). They find  $\log A_{\text{yr}^{-1}} \approx -15 \pm 0.5$  with an turnover in the GWB below  $\sim 10^{-2} \text{ yr}^{-1}$  and significant attenuation up to  $\sim 10^{-1} \text{ yr}^{-1}$ . Recently, both [McWilliams, Ostriker & Pretorius \(2014\)](#) and [Kulier et al. \(2015\)](#) have implemented explicit dynamical-friction formalisms along with recent MBH-Host scaling relations ([McConnell & Ma 2013](#)) applied to halo mass functions from Press-Schechter and the Millennium simulations respectively. [McWilliams, Ostriker & Pretorius \(2014\)](#) find  $\log A_{\text{yr}^{-1}} \approx 14.4 \pm 0.3$ , and [Kulier et al. \(2015\)](#)  $\log A_{\text{yr}^{-1}} \approx 14.7 \pm 0.1$ , with both highlighting the non-negligible fraction of binaries stalled at kiloparsec-scale separations. Almost all previous studies had assumed that all MBHB merge effectively.

These predictions are summarized in [Table 1.1](#). While far from exhaustive, we believe they are a representative sample, with specific attention to recent work on environmental effects. The amplitudes of the predicted backgrounds are distributed fairly consistently around  $A_{\text{yr}^{-1}} \approx 10^{-15}$ . Assuming observational baselines of about 10 yr, pulsar TOA accuracies of at least tens of microseconds are required to constrain or observe a GWB with this amplitude (see, e.g. [Blandford, Romani & Narayan 1984](#); [Rajagopal & Ro-](#)

\*We use the term ‘Semi-Analytic Model’ loosely to refer to a *realized population* constructed by an analytic prescription, as apposed to derived from underlying physical models.

mani 1995). Finding more millisecond pulsars with very small intrinsic timing noise are key to improving GWB upper-limits, while increasing the total number (and angular distribution) of pulsars will be instrumental for detections (Taylor et al. 2016b).

There are currently three ongoing PTA groups, the North-American Nanohertz Observatory for Gravitational-waves (NANOGrav, The NANOGrav Collaboration et al. 2015), the European PTA (EPTA, Desvignes et al. 2016), and the Parkes PTA (PPTA, Manchester et al. 2013). Additionally, the International PTA (IPTA, Hobbs et al. 2010) aims to combine the data sets from each individual project, and has recently produced their first public data release (Verbiest et al. 2016). Table 1.2 summarizes the current upper limits from each PTA. These are the  $2\text{-}\sigma$  upper bounds, based on both extrapolation to  $A_{\text{yr}^{-1}}$  along with that of the specific frequency with the *strongest constraint* assuming a  $-2/3$  spectral index. Overall, the lowest bound is from the PPTA, at  $A_{\text{yr}^{-1}} < 10^{-15}$ , or in terms of the fractional closure density,  $\Omega_{\text{GW}}(f = 0.2 \text{ yr}^{-1}) < 2.3 \times 10^{-10}$  (Shannon et al. 2015).

| Reference                               | GWB Amplitude*            | Populations |             |                           | Spectral Slope            |
|---|---------------------------|-------------|-------------|---------------------------|---------------------------|
|   | $\log A_{\text{yr}^{-1}}$ | Galaxies    | Black Holes | MBHB Evolution            | [Deviations from $-2/3$ ] |
| Jaffe & Backer (2003)                   | -16                       | SAM         | SAM         | GW                        | -                         |
| Wyithe & Loeb (2003)                    | -14.3                     | SAM         | SAM         | GW                        | -                         |
| Kocsis & Sesana (2011)                  | $-15.7 \pm 0.3$           | Cosmo-DM    | SAM         | VD, GW                    | flattened                 |
| Sesana (2013b)                          | $-15.1 \pm 0.3$           | SAM         | SAM         | GW                        | -                         |
| McWilliams, Ostriker & Pretorius (2014) | $-14.4 \pm 0.3$           | SAM         | SAM         | DF, GW                    | imposed cutoff            |
| Ravi et al. (2014)                      | $-14.9 \pm 0.25$          | Cosmo-DM    | SAM         | LC-Full <sup>†</sup> , GW | flattened & cutoff        |
| Kulier et al. (2015)                    | $-14.7 \pm 0.1$           | Cosmo-Hydro | SAM         | DF, GW                    | -                         |
| Roebber et al. (2016)                   | $-15.2^{+0.4}_{-0.2}$     | Cosmo-DM    | SAM         | GW                        | -                         |
| Sesana et al. (2016)                    | $-15.4 \pm 0.4$           | SAM         | SAM         | GW                        | -                         |

\*Some values which were not given explicitly in the included references were estimated based on their figures, and thus should be taken as approximate.

<sup>†</sup>In this case, the LC prescription is effectively always ‘Full’.

**Table 1.1:** Representative sample of previous predictions for the GW Background, with a basic summary of their implementation. ‘Semi-Analytic Model (SAM)’ is used loosely to refer to numerical models based on scaling relations and observed populations. ‘Cosmo-DM’ are cosmological Dark-Matter only (N-Body) simulations, while ‘Cosmo-Hydro’ are hydrodynamic simulations including baryons. The physical evolution effects included are: Dynamical Friction (DF), Loss-Cone (LC) stellar scattering, Viscous Drag (VD) from a circumbinary disk, and Gravitational Wave (GW) radiation. In this study we use populations of both galaxies and MBH which coevolved in the cosmological, hydrodynamic simulations ‘Illustris’, and include all mechanisms of hardening (DF, LC, VD & GW) in our models.

| PTA      | $A_{\text{yr}^{-1}}$  | Strongest Constraint  |                            | Reference                                 |
|----------|-----------------------|-----------------------|----------------------------|---|
|          |                       | $A_{f,0}$             | $f_0$ [ $\text{yr}^{-1}$ ] |   |
| European | $3.0 \times 10^{-15}$ | $1.1 \times 10^{-14}$ | 0.16                       | <a href="#">Lentati et al. (2015)</a>     |
| NANOGrav | $1.5 \times 10^{-15}$ | $4.1 \times 10^{-15}$ | 0.22                       | <a href="#">Arzoumanian et al. (2016)</a> |
| Parkes   | $1.0 \times 10^{-15}$ | $2.9 \times 10^{-15}$ | 0.2                        | <a href="#">Shannon et al. (2015)</a>     |
| IPTA     | $1.5 \times 10^{-15}$ | -                     | -                          | <a href="#">Verbiest et al. (2016)</a>    |

**Table 1.2:** Upper Limits on the GW Background from Pulsar Timing Arrays. Values are given both at the standard normalization of  $f = 1 \text{ yr}^{-1}$  in addition to the frequency and amplitude of the strongest constraint (when given).

Every existing prediction has been made with the use of SAM—mostly in the construction of the galaxy population, but also in how black holes are added onto those galaxies. SAM are extremely effective in efficiently creating large populations based on observational relations. Higher-order, less observationally constrained parameters can have systemic biases however, for example galaxy merger rates (see, e.g. Hopkins et al. 2010)—which are obviously critical to understanding MBHB evolution. Recently, Rodriguez-Gomez et al. (2015) have shown that merger rates from the cosmological, hydrodynamic Illustris simulations (Vogelsberger et al. 2014a) show excellent agreement with observations, while differing (at times substantially) from many canonical SAM.

In this paper, we use results from the Illustris (Genel et al. 2014) simulations (discussed in §1.2) to make predictions for the rates at which MBH form binaries, and evolve to coalescence. Illustris provides the MBH population along with their self-consistently derived parent galaxies and associated stellar, gaseous, and dark-matter components. This is the first time that a hydrodynamic galaxy population, with fully co-evolved MBH have been used to calculate a GWB spectrum. We use these as the starting point for post-processed models of the unresolved merger dynamics themselves, including all of the underlying hardening mechanisms: dynamical friction, loss-cone stellar scattering, viscous drag, and GW evolution—again for the first time (§1.3) in a MBHB population calculation. From these data, we make predictions for plausible GW backgrounds observable by PTA, focusing on the effects of different particular mechanisms on the resulting spectrum such that future detections and upper-limits can be used to constrain the physical merger process (§1.4).

In addition to the GWB, understanding the population of MBHB is also important for future space-based GW observatories (e.g. eLISA The eLISA Consortium et al. 2013). Solid predictions for binary timescales at different separations will also be instrumental in interpreting observations of dual- and binary-AGN, in addition to offset and ‘kicked’ BHs (Blecha et al. 2016). Finally, MBHB could play a significant role in triggering stellar Tidal Disruption Events (TDE; e.g. Ivanov, Polnarev & Saha 2005b; Chen et al. 2009), and explaining the distribution of observed TDE host galaxies.

## 1.2 THE ILLUSTRIS SIMULATIONS

Illustris are a suite of cosmological, hydrodynamic simulations which have accurately reproduced both large-scale statistics of thousands of galaxies at the same time as the detailed internal structures of ellipticals and spirals (Vogelsberger et al. 2014a). Illustris—hereafter referring to Illustris-1, the highest resolution of three runs—is a cosmological box of 106.5 Mpc on a side, with  $1820^3$  each gas cells and dark-matter particles. The simulations use the moving, unstructured-mesh hydrodynamic code AREPO (Springel 2010), with superposed SPH particles (e.g. Springel et al. 2005) representing stars (roughly  $1.3 \times 10^6 M_\odot$  mass resolution, 700 pc gravitational softening length), dark-matter (DM;  $6.3 \times 10^6 M_\odot$ , 1.4 kpc), and MBH (seeded at  $M \approx 10^5 M_\odot$ ) and allowed to accrete and evolve dynamically. Stars form and evolve, feeding back and enriching their local environments, over the course of the simulation which is initialized at redshift  $z = 137$  and evolved until  $z = 0$  at which point there are over  $3 \times 10^8$  star particles.

For a comprehensive presentation of the galaxy formation models (e.g. cooling, inter-stellar medium, stellar evolution, chemical enrichment) see the papers: Vogelsberger et al. (2013) and Torrey et al. (2014). For detailed descriptions of the general results of the Illustris simulations, and comparisons of their properties with the observed universe, see e.g., Vogelsberger et al. (2014b), Genel et al. (2014), and Sijacki et al. (2015). Finally, the data for the Illustris simulations, and auxiliary files containing the black hole data\*

\*The blackhole data files were made public in late Sept. 2016.

used for this analysis, have been made publicly available online ([www.illustris-project.org](http://www.illustris-project.org); Nelson et al. 2015).

### 1.2.1 THE BLACK HOLE MERGER POPULATION

Black holes are implemented as massive, collisionless ‘sink’ particles seeded into sufficiently massive halos. Specifically, halos with a total mass above  $7.1 \times 10^{10} M_{\odot}$ , identified using an on the fly Friends-Of-Friends (FOF) algorithm, which don’t already have a MBH are given one with a seed mass,  $M_{\text{seed}} = 1.42 \times 10^5 M_{\odot}$  (Sijacki et al. 2007). The highest density gas cell in the halo is converted into the BH particle. The BH mass is tracked as an internal quantity, while the particle overall retains a *dynamical* mass initially equal to the total mass of its predecessor gas cell (Vogelsberger et al. 2013). The internal BH mass grows by Eddington-limited, Bondi-Hoyle accretion from its parent gas cell (i.e. the total dynamical mass remains the same). Once the excess mass of the parent is depleted, mass is accreted from nearby gas particles—increasing both the dynamical mass of the sink particle, and the internal BH-mass quantity.

BH sink particles typically have masses comparable to (or within a few orders of magnitude of) that of the nearby stellar and DM particles. Freely evolving BH particles would then scatter around their host halo, instead of dynamically settling to their center—as is the case physically. To resolve this issue, BH particles in Illustris are repositioned to the potential minima of their host halos. For this reason, their parametric velocities are not physically meaningful. Black hole “mergers” occur in the simulation whenever two MBH particles come within a particle smoothing-length of one another—typically on the order of a kiloparsec. This project aims to fill in the merger process unresolved in Illustris. In our model, an Illustris “merger” corresponds to the *formation* of a MBH binary system, which we then evolve. To avoid confusion, we try to use the term ‘coalescence’ to refer to the point at which such a binary would actually collide, given arbitrary resolution.

Over the course of the Illustris run, 135 ‘snapshots’ were produced, each of which include internal parameters of all simulation particles. Additional black-hole-specific output was also recorded at every time-step, providing much higher time resolution for black hole accretion rates\*, local gas densities and most notably, merger events. The entire set of mergers—a time and pair of BH masses—constitutes our initial population of MBH binaries.

The distribution of BH masses is peaked at the lowest masses. Many of these black holes are short-lived: their small, usually satellite, host of-matter halos often quickly merge with a nearby neighbor—producing a BH ‘merger’ event. Additionally, in some cases, the identification of a particular matter overdensity as a halo by the FOF halo-finder, while transient, may be sufficiently massive to trigger the creation of a new MBH seed particle. This seed can then quickly merge with the MBH in a nearby massive halo. Due to the significant uncertainties in our understanding of MBH and MBH-seed formation, it is unclear if and when these processes are physical. For this reason we implement a mass cut on merger events, to ensure that each component BH has  $M_{\bullet} > 10^6 M_{\odot} \approx 10 M_{\text{seed}}$ . Whether or not these ‘fast-mergers’ are non-physical, the mass cut is effective at excluding them from our analysis. The entire Illustris simulation has 23708 MBH merger events; applying the mass cut excludes 11291 (48%) of those, leaving 12417. We have run configurations without this mass cut, and the effects on the GWB are always negligible.

There is very small population of MBH ‘merger’ events which occur during close encounters (but not true mergers) of two host-halos. During the encounter, the halo-finder might associate the two constituent halos as one, causing the MBHs to merge spuriously due to the repositioning algorithm. These

\*These, self-consistently derived mass accretion rates ( $\dot{M}$ ) are used in our implementation of gas drag (discussed in §1.3.3) as a way of measuring the local gas density.

forced mergers are rare and, we believe, have no noticeable effect on the overall population of thousands of mergers. They are certainly negligible in the overall MBH-Halo statistical correlations which are well reproduced in the Illustris simulations (Sijacki et al. 2015).

### 1.2.2 MERGER HOST GALAXIES

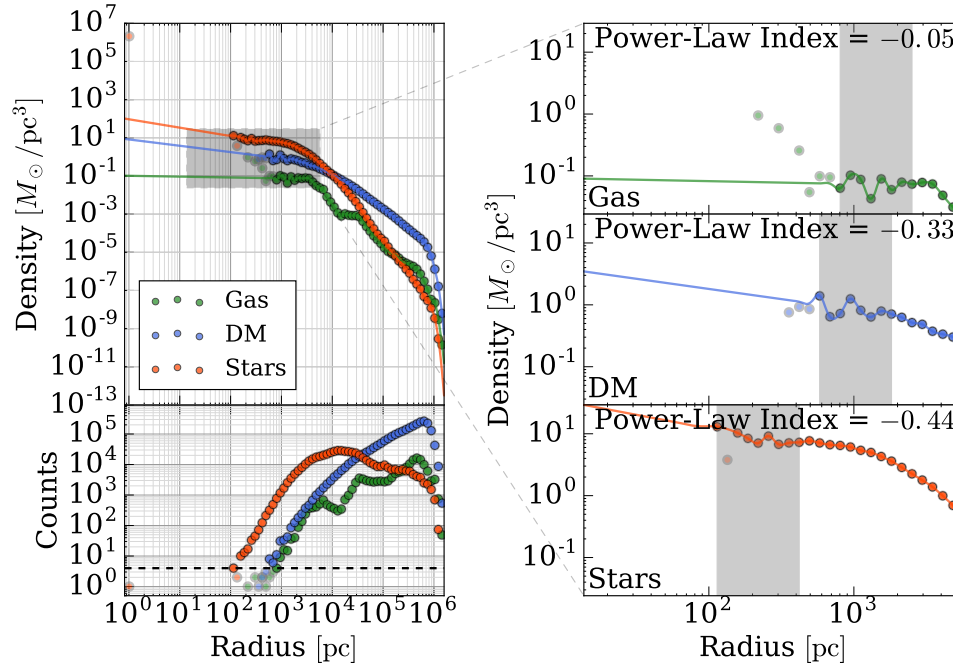
To identify the environments which produce the dynamical friction, stellar scattering, and viscous drag which we are interested in, we identify the host galaxies of each MBH involved in the merger in the snapshot preceding it, in addition to the single galaxy which contains the ‘binary’ (at this point a single, remnant MBH) in the snapshot immediately following the merger event. 644 mergers (3%) are excluded because they don’t have an associated galaxy before or after the merger. To ensure that each host galaxy is sufficiently well resolved (especially important for calculating density profiles), we require that a sufficient number of each particle type constitute the galaxy. Following Blecha et al. (2016) we use a fiducial cut of 80 and 300 star and DM particles respectively, and additionally require 80 gas cells. This excludes 54 of the remaining binaries. We emphasize here that this remnant host galaxy, as it is in the snapshot following the Illustris ‘merger’ event, forms the environment in which we model the MBHB merger process. In the future, we plan to upgrade our implementation to take full advantage of the dynamically evolving merger environment: including information from both galaxies as they merge, and the evolving remnant galaxy once it forms.

We construct spherically averaged, radial density profiles for each host galaxy and each particle/cell type (star, DM, gas). Because the particle smoothing lengths are larger than the MBHB separations of interest, we extrapolate the galaxy density profiles based on fits to the inner regions. Our fits use the innermost eight radial bins that have at least four particles in them. Out of the valid binaries, 347 (1%) are excluded because fits could not be constructed—generally because the particles aren’t distributed over the required eight bins. Successful fits typically use  $\sim 100$  particles, with gas cell sizes  $\sim 10^2$  pc and SPH smoothing lengths for stars and DM  $\sim 10^3$  pc.

Density profiles for a sample Illustris MBHB host galaxy are shown in Fig. 1.1. The left panels show the binned density profiles for each particle type (top), and the number of particles in each bin (bottom). The semi-transparent points are those with less than the requisite four particles in them. The right panels show zoom-ins for each particle type, where the shaded regions indicate the eight bins used for calculating fits. The resulting interpolants are overplotted, with the power-law index indicated. While the four particles per bin, and eight inner bins generally provide for robust fits, we impose a maximum power-law index of  $-0.1$ —i.e. that densities are at least gently increasing, and a minimum index of  $-3$ —to ensure that the mass enclosed is convergent. Using these densities we calculate all additional galaxy profiles required for the hardening prescriptions (§1.3), e.g. velocities, binding energies, etc. When calculating profiles using our fiducial parameters, 2286 binaries (10%) are excluded when calculating the distribution functions (§1.3.2), usually due to significant nonmonotonicities in the radial density profile which are incompatible with the model assumptions. Overall, after all selection cuts, 9270/23708 (39%) of the initial Illustris ‘merger’ events are analyzed in our simulations.

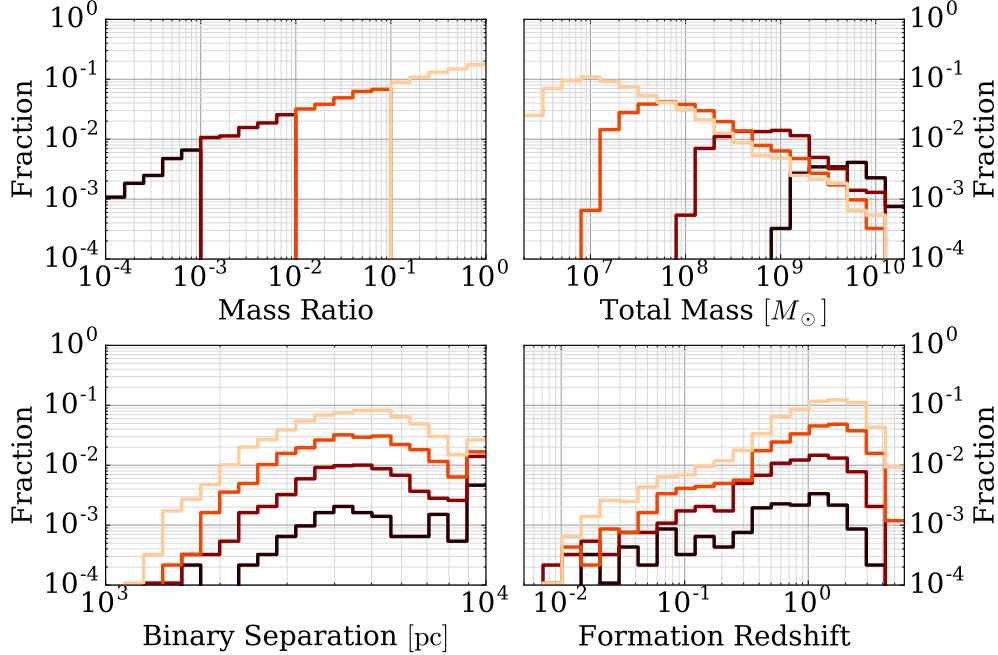
Figure 1.2 shows the properties of MBHB passing our selection cuts, grouped by mass ratio. Mass ratio (upper-left panel) is strongly anti-correlated with total mass (upper-right) due to both selection effects (e.g. at total masses just above the minimum mass, the mass ratio must be near unity), and astrophysical ones (e.g. the most massive MBH, in large, central galaxies tend to merge more often with the lower mass MBH in small satellite galaxies). Binary separations (lower-left) are set by the smoothing length of MBH particles in Illustris. Once two MBH particles come within a smoothing length of one





**Figure 1.1:** Density profiles from a sample Illustris MBHB host galaxy. Binned densities for each particle type are shown (upper-left) along with the number of particles/cells in each (bottom-left). Semi-transparent points are bins with less than four particles—the number required for consideration in calculating fits. Zoom-ins are also shown separately for each particle type (right), with the eight inner-most bins with at least four particles shown in the shaded region. Those bins were used to calculate fits, which are overplotted. The resulting power-law indices used to extrapolate inwards are also shown. Any galaxy without enough (eight) bins is excluded from our sample, in addition to the MBHB it contains. This was the case for roughly 1% of our initial population, almost entirely containing MBH very near our BH mass threshold ( $10^6 M_{\odot}$ ).

another, Illustris considers them a ‘merger’ event—which corresponds to the ‘formation’ (lower-right) of a binary in the simulations of this study.



**Figure 1.2:** Properties of the Massive Black Hole Binaries from the Illustris population passing our selection cuts. After selecting for MBH masses  $M > 10^6 M_\odot$ , and requiring the binary host galaxies to have sufficiently well resolved density profiles, 9270 of 23708 (39%) systems remain. Distributions of mass ratio, total mass, initial binary separation (determined by MBH particle smoothing lengths), and formation redshifts (determined as the time at which particles come within a smoothing length of one-another) are shown. The different lines (colors) correspond to different mass ratios which are strongly anti-correlated with total mass.

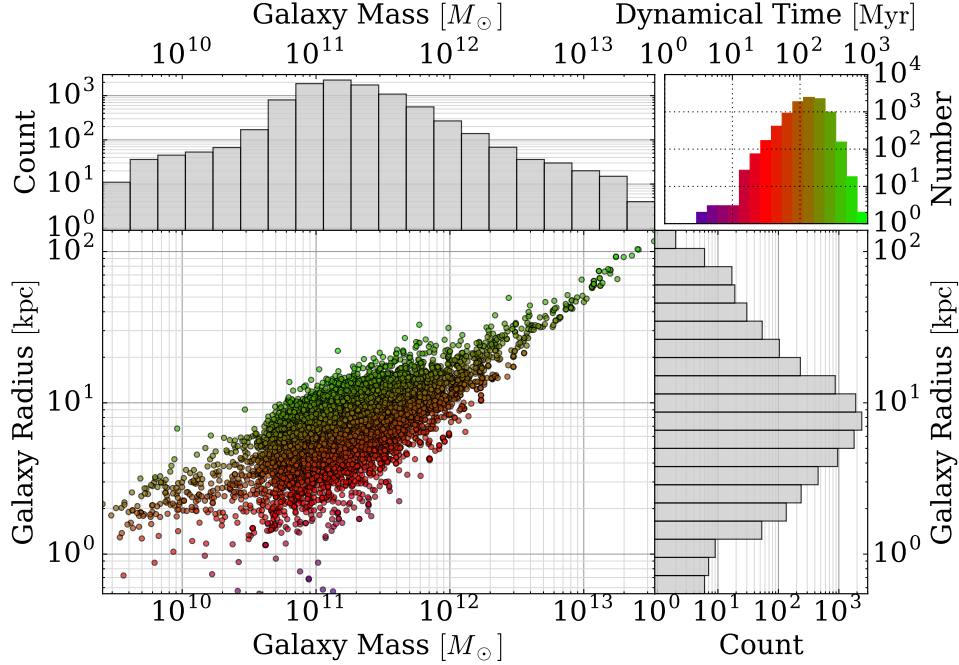
### 1.3 BINARY HARDENING MODELS

Black-hole encounters from Illustris determine the initial conditions for the binary population which are then evolved in our merger simulations. Throughout the ‘hardening’ process, where the binaries slowly coalesce over millions to billions of years, we assume uniformly circular orbits. In our models, we use information from the MBHB host galaxies to implement four distinct hardening mechanisms (Begelman, Blandford & Rees 1980, hereafter BBR80): dynamical friction (DF), stellar scattering in the ‘loss-cone’ regime (LC)\*, viscous drag (VD) from a circumbinary gaseous-disk, and gravitational-wave radiation (GW).

#### 1.3.1 DYNAMICAL FRICTION

Dynamical friction is the integrated effect of many weak and long-range scattering events, on a gravitating object moving with a relative velocity through a massive background. The velocity differential causes an asymmetry which allows energy to be transferred from the motion of the massive object to a kinetic thermalization of the background population. In the case of galaxy mergers, dynamical friction is the primary mechanism of dissipating the initial orbital energy to facilitate coalescence of the galaxies, generally on timescales comparable to the local dynamical time ( $\sim 10^8$  yr). BH present in the parent galaxies will tend to ‘sink’ towards each other in the same manner (BBR80) due to the background of stars, gas

\*Loss-cone scattering and dynamical friction are different regimes of the same phenomenon, we separate them based on implementation.



**Figure 1.3:** Stellar half-mass radii ( $R_{*,1/2}$ ) and total mass within  $2 \cdot R_{*,1/2}$  for all Illustris remnant host galaxies. Values for each galaxy are colored by their dynamical times, calculated using Eq. 1.7, which are used for the ‘enhanced’ DF masses. The histogram in the upper right shows the distribution of dynamical times. Galaxy masses and radii are peaked at about  $10^{11} M_{\odot}$  and 10 kpc respectively, with corresponding dynamical times around 100 Myr

and dark-matter (DM). For a detailed review of dynamical friction in MBH systems, see [Antonini & Merritt \(2012\)](#).

The change in velocity of a massive object due to a single encounter with a background particle at a fixed relative velocity  $v$  and impact parameter  $b$  is derived (e.g. [Binney & Tremaine 1987](#)) following the treatment of [Chandrasekhar \(1942, 1943\)](#) by averaging the encounters over all possible angles to find,

$$\Delta v = -2v \frac{m}{M+m} \frac{1}{1+(b/b_0)^2}, \quad (1.3)$$

where the characteristic (or ‘minimum’) impact parameter  $b_0 \equiv G(M+m)/v^2$ , for a primary object of mass  $M$ , in a background of bodies with masses  $m$ . The net deceleration on a primary mass is then found by integrating over distributions of stellar velocity (assumed to be isotropic and Maxwellian) and impact parameters (out to some maximum effective distance  $b_{\max}$ ) which yields,

$$\frac{dv}{dt} = -\frac{2\pi G^2(M+m)\rho}{v^2} \ln \left[ 1 + (b_{\max}/b_0)^2 \right], \quad (1.4)$$

for a background of mass density  $\rho$ . The impact parameters are usually replaced with a constant—the ‘Coulomb Logarithm’,  $\ln \Lambda \equiv \ln \left( \frac{b_{\max}}{b_0} \right) \approx \frac{1}{2} \ln (1 + \Lambda^2)$ , such that,

$$\frac{dv}{dt} = -\frac{2\pi G^2(M+m)\rho}{v^2} \ln \Lambda. \quad (1.5)$$

In the implementation of Eq. 1.5, we use spherically averaged density profiles from the Illustris, remnant host galaxies. Modeling a ‘bare’, secondary MBH moving under DF through these remnants would clearly drastically underestimate the effective mass—which, at early times is the MBH secondary in addition to its host galaxy. Over time, the secondary galaxy will be stripped by tidal forces and drag, eventually leaving behind the secondary MBH with only a dense core of stars and gas directly within its sphere

of influence. We model this mass ‘enhancement’ by assuming that the effective DF mass is initially the sum of the MBH mass ( $M_2$ ), and that of its host galaxy ( $M_{2,\text{host}}$ ), decreasing as a power-law over a dynamical time  $\tau_{\text{dyn}}$ , until only the MBH mass is left, i.e.,

$$M_{\text{DF}} = M_2 \left( \frac{M_2 + M_{2,\text{host}}}{M_2} \right)^{1-t/\tau_{\text{dyn}}} . \quad (1.6)$$

We calculate the dynamical time using a mass and radius from the remnant host galaxy. Specifically, we use twice the stellar half-mass radius  $2R_{\star,1/2}$ , and define  $M_{2,\text{host}}$  as the total mass within that radius, i.e.,

$$\tau_{\text{dyn}} = \left[ \frac{4\pi (2R_{\star,1/2})^3}{3GM_{2,\text{host}}} \right]^{1/2} . \quad (1.7)$$

The galaxy properties and derived dynamical times for all MBHB host galaxies we consider are shown in Fig. 1.3. Galaxy masses and radii are peaked at about  $10^{11} M_{\odot}$  and 10 kpc respectively, with corresponding dynamical times around 100 Myr. For comparison, we also perform simulations using a fixed dynamical time of 1 Gyr for all galaxies, i.e., less-efficient stripping of the secondary galaxy.

### 1.3.1.1 IMPACT PARAMETERS AND EXPLICIT CALCULATIONS

We have explored calculating the Coulomb logarithm explicitly, following BBR80 for the maximum impact parameter such that,

$$b_{\text{max}}(r) = \begin{cases} R_s & R_s < r, \\ (r/R_b)^{3/2} R_s & R_h < r < R_s, \\ r R_h & r < R_h. \end{cases} \quad (1.8)$$

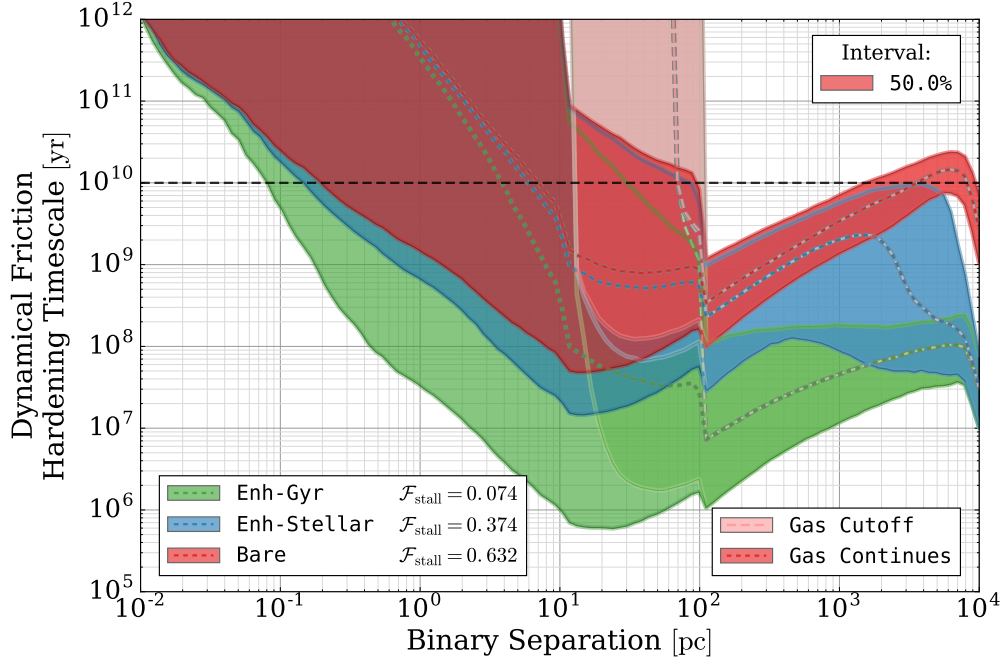
This, effective maximum impact parameter is a function of binary separation\*  $r$ —to account for the varying population of stars available for scattering and varying effectiveness of encounters. Eq. 1.8 also depends on the characteristic stellar radius  $R_s$  ( $‘r_c’$  in BBR80), radius at which the binary becomes gravitationally bound,  $R_b = [M/(Nm_{\star})]^{1/3} R_s$ , and radius at which the binary becomes ‘hard’,  $R_h \equiv (R_b/R_s)^3 R_s$ .

Not only is this formalism complex, but it often produces unphysical results. For example, with this prescription the ‘maximum’ impact parameter not infrequently becomes less than the ‘minimum’, or larger than the distances which interact in the characteristic timescales. After imposing a minimum impact parameter ratio of  $b_{\text{max}}/b_0 \geq 10$  (i.e.  $\ln \Lambda \geq 2.3$ ), the results we obtained are generally consistent with using a constant coulomb-logarithm, with negligible effects on the resulting merger rates and GW background. We have also implemented an explicit integration over stellar distribution functions (see: §1.3.2), and found the results to again be entirely consistent with Eq. 1.5 which is both computationally faster and numerically smoother. We believe the explicit impact parameter calculation is only valuable as a heuristic, and instead we use  $\ln \Lambda = 15$ , consistent with detailed calculations (e.g. Antonini & Merritt 2012). Similarly, in the results we present, we take the local stellar density as that given by spherically symmetric radial density profiles around the galactic center instead of first determining, then marginalizing over, the stellar distribution functions.

### 1.3.1.2 APPLICABLE REGIMES

There is a critical separation at which the back-reaction of the decelerating MBH notably modifies the stellar distribution, and the dynamical friction formalism is no longer appropriate. Beyond this radius, the finite number of stars in the accessible region of parameter space to interact with the MBH(B)—the

\*Note that we use the term ‘binary separation’ loosely, in describing the separation of the two MBH even before they are gravitationally bound.



**Figure 1.4:** Dynamical Friction (DF) hardening timescales for 50% of our MBHB around the median, under a variety of implementations. Cases in which a ‘bare’ secondary MBH migrates through the remnant host galaxy is compared to ones where the effective mass is *enhanced* to the secondary’s host galaxy, decreasing as a power-law over the course of a dynamical time (‘Enh’; see Eq. 1.6). The dynamical time is calculated using twice the ‘Stellar’ half-mass radius (see Eq. 1.7) shown in blue, or using a fixed 1 ‘Gyr’ timescale shown in green. Allowing gaseous DF to continue below the attenuation radius  $\mathcal{R}_{lc}$  (‘Gas Continues’: darker regions, dotted lines) is compared to cutting off the gas along with stars and DM (‘Gas Cutoff’: lighter regions, dashed lines).  $\mathcal{F}_{\text{stall}}$  is the fraction of mergers with mass ratio  $\mu > 0.1$ , remaining at separations  $r > 10^2$  pc, at redshift zero.

‘loss-cone’, (LC; see, Merritt 2013)—must be considered explicitly, discussed more thoroughly in §1.3.2.

The ‘loss-cone’ radius can be approximated as (BBR80),

$$\mathcal{R}_{lc} = \left(\frac{m_\star}{M}\right)^{1/4} \left(\frac{R_b}{R_s}\right)^{9/4} R_s. \quad (1.9)$$

Stars and DM are effectively collisionless, so they can only refill the LC on a slow, diffusive scattering timescale. Gas, on the other hand, is viscous and supported thermally and by turbulent motion which can equilibrate it on shorter timescales. In our fiducial model, we assume that for separations  $r < \mathcal{R}_{lc}$  the dynamical friction due to stars and DM is attenuated to low values, but that of gas continues down to smaller separations. We set the inner edge of gaseous DF based on the formation of a (circumbinary) accretion disk on small scales (discussed in §1.3.3). The attenuation prescription given by BBR80 increases the dynamical friction timescales by a factor,

$$f_{\text{DF,LC}} = \left(\frac{m}{M_\bullet}\right)^{7/4} N_\star \left(\frac{R_b}{R_s}\right)^{27/4} \left(\frac{\mathcal{R}_{lc}}{r}\right), \quad (1.10)$$

where  $N_\star = \frac{1}{M_\star} \int_0^r 4\pi r'^2 \rho_\star dr'$ , is the number of stars available to interact with the binary. For all intents and purposes this negates the effectiveness of DF for  $r \lesssim \mathcal{R}_{lc}$ , such that without other hardening mechanisms (which become important at smaller scales), no MBHB would coalesce within a Hubble time.

### 1.3.1.3 DF HARDENING RATES

The resulting hardening timescales,  $\tau_h = a/(da/dt)$ , for our different DF implementations are shown in Fig. 1.4. We show evolution for ‘bare’ MBH secondaries (red), in addition to effective masses enhanced (‘enh’) by the secondary’s host galaxy for a dynamical time calculated using twice the ‘stellar’ half-mass radii (blue), or with a fixed ‘Gyr’ time scale (green). In each case we also compare between letting gas-DF continue below  $\mathcal{R}_{ic}$  (‘Gas Continues’: darker colors, dotted lines), versus attenuating gas along with stars and DM (‘Gas Cutoff’: lighter colors, dashed lines). As a metric of the varying outcomes, we calculate the fraction of ‘stalled’ major-mergers  $\mathcal{F}_{\text{stall}}$ , defined as the number of major mergers (mass-ratios  $\mu \equiv M_2/M_1 > 0.1$ ; which are 6040 out of the full sample of 9270, i.e.  $\sim 65\%$ ) remaining at separations larger than 100 pc at redshift zero, divided by the total number of major-mergers. Attenuation of the DF begins below 100 pc, so  $\mathcal{F}_{\text{stall}}$  are unaffected by whether the gas DF is also cutoff.

For  $r \gtrsim 100$  pc, where the density of stars and especially dark matter dominate that of gas, the hardening rates are the same with or without a separate treatment of gas. Hardening differs significantly however, between the ‘bare’ and enhanced models—with the latter hardening more than an order of magnitude faster at the largest separations\* ( $\sim 10^4$  pc). After a dynamical time, the ‘stellar’ enhancement runs out and the hardening rate approaches that of a bare MBH secondary by  $\sim 10^3$  pc. Still, the enhanced mass over this time leads to a decrease in the fraction of stalled binaries from  $\sim 63\%$  to  $\sim 37\%$ . When the mass enhancement persists for a gigayear—about a factor of ten longer than typical dynamical times—a large fraction of MBHB are able to reach parsec-scale separations before tidal stripping becomes complete, leading to only  $\sim 7\%$  of major-mergers stalling at large separations. The particular fraction of stalled systems is fairly sensitive to the total mass and mass ratio cutoff, which we return to in §1.4.3.

Previous studies (e.g. Ravi et al. 2014) have assumed that DF is very effective at bringing MBHB into the dynamically ‘hard’ regime (i.e. instantly in their models), after which stellar interactions must be calculated explicitly to model the remaining evolution. For comparison, in our results we also include a ‘Force-Hard’ model in which we assume that all binaries reach the hard regime ( $r = R_h$ ) over the course of a dynamical time<sup>†</sup>.

It has long been suggested that MBHB could stall at kiloparsec-scale separations (e.g. Yu 2002), but only recently has this effect been incorporated into population hardening models and studied specifically (e.g. McWilliams, Ostriker & Pretorius 2014; Kulier et al. 2015). By better understanding the timescales over which stalled systems could be observable, we can use observed dual-AGN to constrain the hardening process and the event rates of MBHB encounters.

### 1.3.2 STELLAR LOSS-CONE SCATTERING

The population of stars that are able to interact (scatter) with the MBHB are said to occupy the ‘loss-cone’ (LC, Merritt 2013), so-named because it describes a conical region in parameter space. When stars are scattered out of the LC faster than they can be replenished, the LC becomes depleted and the dynamical friction description, which considers a relatively static background, becomes inconsistent. One approach to compensate for this is to add an ‘attenuation’ factor, as described in the previous section (§1.3.1). Physically, a steady state must be dynamically realized in which stars are diffused into the outer edges of the LC via two-body relaxation at the same rate at which stars are scattering out by the central

\*Recall that binary separations are initialized to the MBH particle smoothing lengths, distributed between about  $10^3$  and  $10^4$  pc, so the total number of systems being plotted decreases over the same range.

<sup>†</sup>calculated using the ‘Stellar’ prescription: twice the stellar half mass radius, and the mass within it.

MBH(B). The largest uncertainty in the MBHB merger process is likely understanding the nature of this equilibrium state, and how it is affected by realistic galaxy-merger environments.

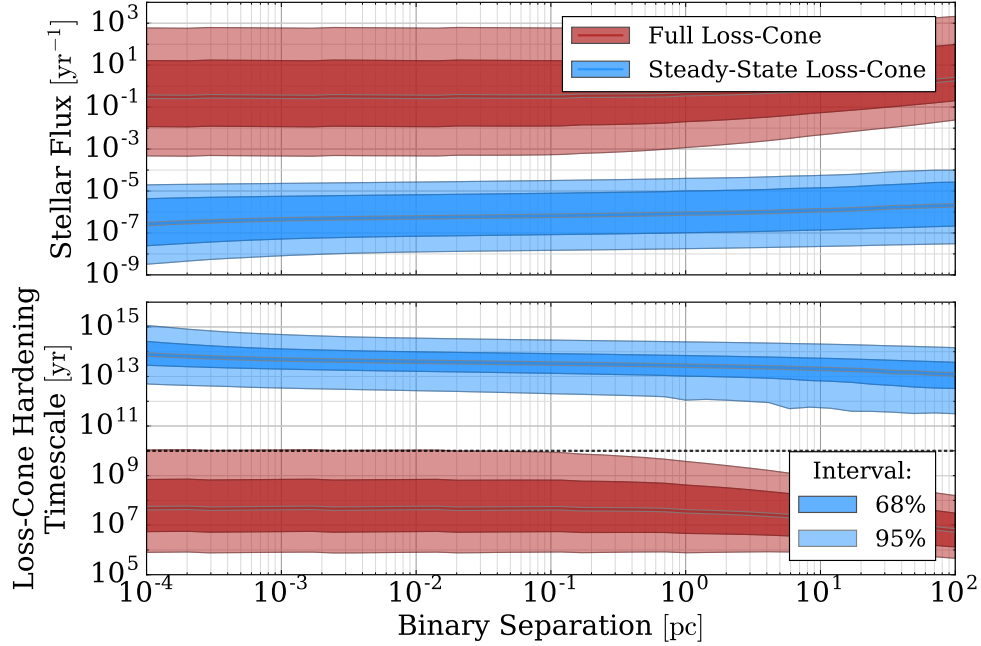
The loss-cone has been extensively explored in both the context of MBH binary hardening, as well as in tidal-disruption events (TDE). The two cases are almost identical, differing primarily in that for TDE calculations, only impact parameters small enough to cause disruptions are of interest—while for binary hardening, weaker scattering events are still able to extract energy from the MBH or MBHB. For binary hardening, there is an additional ambiguity in two subtly distinct regimes: first, where stars scatter with individual BHs, decelerating them analogously to the case of dynamical friction (but requiring the LC population to be considered explicitly). Second, for a truly *bound* binary, stars can interact with the combined system—in a three-body scattering—and extract energy from the binary pair together. In our prescriptions we do not distinguish between these cases, considering them to be a spectrum of the same phenomenon instead.

We use the model for LC scattering given by [Magorrian & Tremaine \(1999\)](#), corresponding to a single central object in a spherical (isotropic) background of stars. We adapt this prescription simply by modifying the radius of interaction to be appropriate for scattering with a binary instead of being tidally disrupted by a single MBH. This implementation is presented in pedagogical detail in Appendix §A. Scattering rates are calculated corresponding to both a ‘full’ LC (Eq. 1.24), one in which it is assumed that the parameter space of stars is replenished as fast as it is scattered; in addition to a ‘steady-state’ LC (Eq. 1.25), in which diffusive two-body scattering sets the rate at which stars are available to interact with the binary.

The interaction rates (fluxes) of stars scattering against all MBHB in our sample are shown in the upper panel of Fig. 1.5 for both full (red, Eq. 1.24) and equilibrium (blue, Eq. 1.25) loss-cone configurations. The interaction rates for full LC tend to be about six orders of magnitude higher than equilibrium configurations. The resulting binary hardening timescales are shown in the lower panel of Fig. 1.5—reaching four orders of magnitude above and below a Hubble time. Clearly, whether the LC is in the relatively low equilibrium state or is more effectively refilled has huge consequences for the number of binaries which are able to coalesce within a Hubble time.

Many factors exist which may contribute to quickly refilling the loss cone. In general, any form of asymmetry in the potential will act as an additional perturber—increasing the thermalization of stellar orbits. The presence of a MBHB is premised on there having been a recent galaxy merger—implying that significant asymmetries and aspherical morphologies may exist. Even ignoring galaxy mergers, galaxies themselves are triaxial (e.g. [Illingworth 1977](#); [Leach 1981](#)), many have bars (e.g. [Sellwood & Wilkinson 1993](#)), and in star forming galaxies there are likely large, dense molecular clouds (e.g. [Young & Scoville 1991](#)) which could act as perturbers. Finally, because binary lifetimes tend to be on the order of the Hubble time while galaxies typically undergo numerous merger events (e.g. [Rodriguez-Gomez et al. 2015](#)), subsequent merger events can lead to triple MBH systems (see: §1.4.4) which could be very effective at stirring the stellar distribution. While there is some evidence that for galaxy-merger remnants the hardening rate can be nearly that of a ‘full’ LC (e.g. [Khan, Just & Merritt 2011](#)), the community seems to be far from a consensus (e.g. [Vasiliev, Antonini & Merritt 2014](#)), and a purely numerical solution to the LC problem is currently still unfeasible.

In the future, we plan on incorporating the effects of triaxiality and tertiary MBH to explore self-consistent LC refilling. In our current models, we introduce an arbitrary dimensionless parameter—the logarithmic ‘refilling fraction’  $\mathcal{F}_{\text{refill}}$  (in practice, but not requisitely, between  $[0.0, 1.0]$ )—to logarithmically



**Figure 1.5:** Scattering rates and hardening timescales for full (red) and steady-state (blue) loss-cones. The bands represent 68% and 95% of the population around the median. The difference between the two extremes of LC states is a stark six orders of magnitude, illustrating how strong of an effect the LC can have on MBHB mergers. We use a simple, single parameter prescription to describe the state of the LC: the fraction, in log-space, between steady-state and full,  $\mathcal{F}_{\text{refill}}$  (see, Eq. 1.11).

interpolate between the fluxes of steady-state ( $F_{\text{lc}}^{\text{eq}}$ ) and full LC ( $F_{\text{lc}}^{\text{full}}$ ), i.e.,

$$F_{\text{lc}} = F_{\text{lc}}^{\text{eq}} \cdot \left( \frac{F_{\text{lc}}^{\text{full}}}{F_{\text{lc}}^{\text{eq}}} \right)^{\mathcal{F}_{\text{refill}}}. \quad (1.11)$$

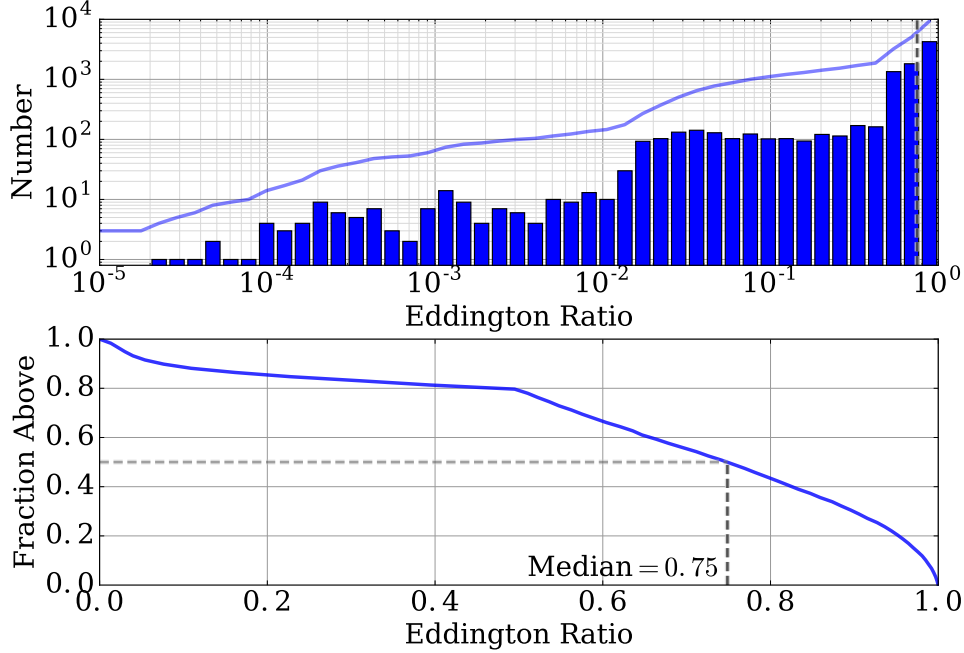
### 1.3.3 VISCOUS HARDENING BY A CIRCUMBINARY DISK

The density of gas accreting onto MBH can increase significantly at separations near the accretion or Bondi radius,  $R_b \equiv GM_{\bullet}/c_s^2$ , where the sound-crossing time is comparable to the dynamical time. The nature of accretion flows onto MBH near and within the Bondi-radius are highly uncertain, as observations of these regions are currently rarely resolved (e.g. Wong et al. 2011). If a high density, circumbinary disk is able to form, the viscous drag (VD) can be a significant contribution to hardening the binary at separations just beyond the GW-dominated regime (BBR80, Gould & Rix 2000; Escala et al. 2005b). Galaxy mergers are effective at driving significant masses of gas to the central regions of post-merger galaxies (Barnes & Hernquist 1992), enhancing this possibility.

We implement a prescription for VD due to a circumbinary accretion disk following Haiman, Kocsis & Menou (2009, hereafter HKM09) based on the classic thin-disk solution of Shakura & Sunyaev (1973), broken down into three, physically distinct regions (Shapiro & Teukolsky 1986). These regions are based on the dominant pressure (radiation vs. thermal) and opacity (Thomson vs. free-free) contributions, such that the regions are defined as,

- 1)  $r < r_{12}$ , radiation pressure and Thomson-scattering opacity dominated;
- 2)  $r_{12} < r < r_{23}$ , thermal pressure and Thomson-scattering opacity dominated;





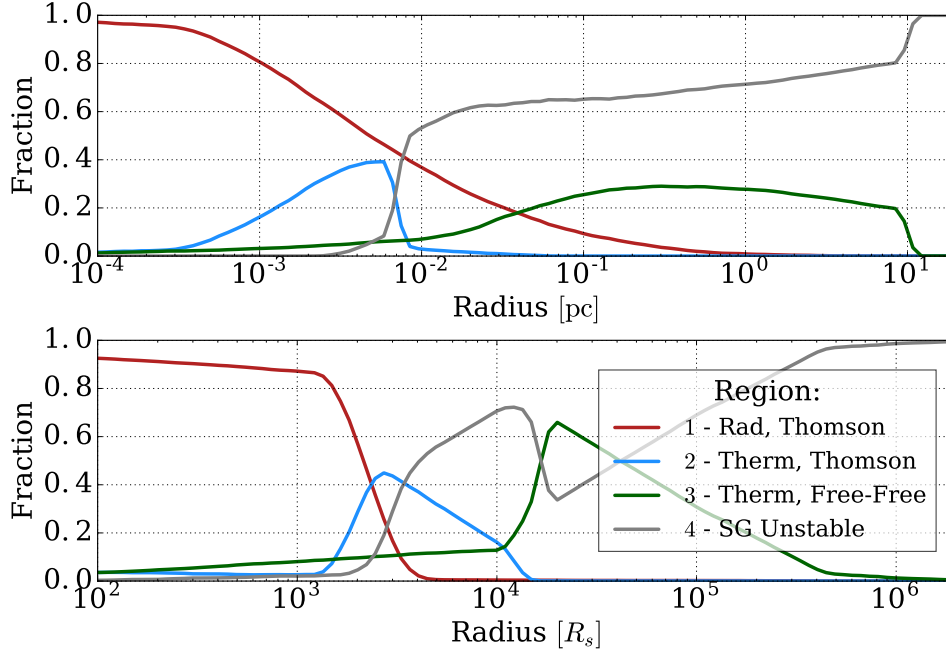
**Figure 1.6:** Accretion rates at the time of binary formation for all MBHB in our analysis. Values are measured as a fraction of the Eddington accretion rate,  $\dot{M}_{\text{Edd}} \equiv L_{\text{Edd}}/\varepsilon_{\text{rad}}c^2$ , where we use  $\varepsilon_{\text{rad}} = 0.1$ . Recall that in Illustris MBH merge when they come within a smoothing length of one another—corresponding to the formation of a binary in our models. The accretion rates from Illustris are those of the resulting remnant MBH, and are limited to Eddington ratios of unity. The upper panel shows the distribution of accretion rates (bars) and cumulative number (line), which are strongly biased towards near-Eddington values. The lower panel shows the cumulative distribution of accretion rates above each value (note the different x-axis scaling). The median Eddington ratio of 0.75 is overplotted (grey, dashed line).

3)  $r_{23} < r$ , thermal pressure and free-free opacity dominated.

Recall that in Illustris, ‘mergers’ occur when MBH particles come nearer than a particle smoothing length, after which the MBH are combined into a single, remnant MBH. We track these remnant particles and use their accretion rates ( $\dot{M}$ ) to calibrate the circumbinary disk’s gas density. The distribution of Eddington ratios ( $\dot{M}/\dot{M}_{\text{Edd}}$ ) for these remnants, at the time of their formation, are presented in Fig. 1.6, showing a clear bias towards near-Eddington accretion rates. MBH remnants tend to have enhanced accretion rates for a few gigayear after merger, and have higher average accretion than general BHs (Blecha et al. 2016). In Illustris, MBH accretion (and thus growth in mass) is always limited to the Eddington accretion rate. We introduce a dimensionless parameter  $f_{\text{Edd}}$  to modulate those accretion rates, i.e.,  $\dot{M} = \text{Min} [\dot{M}_{\text{ill}}, f_{\text{Edd}}\dot{M}_{\text{Edd}}]$ .

Otherwise, in the formalism of HKM09, we use their fiducial parameter values\* and assume an  $\alpha$ -disk (i.e. the viscosity depends on total pressure, not just thermal pressure as in a so-called  $\beta$ -disk). The outer disk boundary is determined by instability due to self-gravity—measured as some factor times the radius,  $r_Q$ , at which the Toomre parameter reaches unity, i.e.  $\mathcal{R}_{\text{SG}} = \lambda_{\text{sg}} r_Q$ . In our fiducial model,  $\lambda_{\text{sg}} = 1$ , and variations in this parameter have little effect on the overall population of binaries. After marginalizing over all systems, changes to the different viscous-disk parameters tend to be largely degenerate: shifting

\*Mean mass per electron,  $\mu_e = 0.875$ ; viscosity parameter  $\alpha = 0.3$ ; radiative efficiency,  $\varepsilon_{\text{rad}} = 0.1$ ; temperature-opacity constant,  $f_T = 0.75$ ; and disk-gap size,  $\lambda_{\text{gap}} = 1.0$  (‘ $\lambda$ ’ in HKM09).



**Figure 1.7:** Fraction of binaries in each circumbinary disk region as a function of radius. Radii are given both physical units (upper panel) and Schwarzschild radii ( $R_s$ , lower panel), the latter highlighting the intrinsic scalings. *Region-4* are locations in the disk which are unstable to self-gravity (‘SG’), defined using the Toomre parameter for each of *Region-2* and *Region-3*.

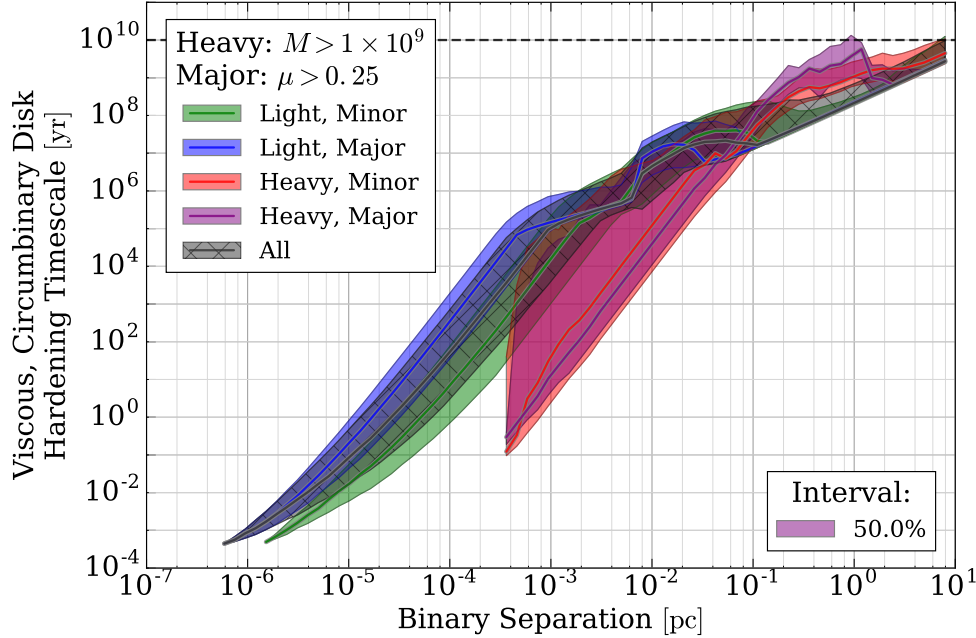
the distribution of hardening timescales and the GWB amplitude in similar manners.

VD hardening timescales tend to decrease with decreasing binary separations. They thus tend to be dominated by *Region-3*—at larger separations. For near-Eddington accretion rates, however, *Region-2* and especially *Region-3* tend to be self-gravity unstable, fragmenting the disk and eliminating VD altogether. For this reason, when high accretion rate systems have dynamically important disks, they tend to be in *Region-1*. In these cases, *Region-1* extends to large enough radii such that for most masses of interest, GW emission will only become significant well within that region of the disk. Lower accretion rate systems are stable out to much larger radii, allowing many binaries to stably evolve through *Region-2* and *Region-3*. These regions also cutoff at smaller separations, meaning that GW emission can become significant outside of *Region-1*.

Decreased disk densities mean less drag, but at the same time sufficiently high densities lead to instability, making the connection between accretion rate and VD-effectiveness non-monotonic. This is enhanced by gaseous DF, with an inner cutoff radius determined by the SG radius (see §1.3.1.2). In other words, gaseous DF is allowed to continue down to smaller radii when the outer disk regions become SG unstable. We impose an additional, absolute upper-limit to the SG instability radius of  $\mathcal{R}_{\text{SG,Max}} = 10$  pc, i.e.  $\mathcal{R}_{\text{SG}} = \text{Min}[\lambda_{\text{sg}} r_Q, \mathcal{R}_{\text{SG,Max}}]$ , to keep the outer-edge of disks physically reasonable.

Fig. 1.7 shows the fraction of Illustris binaries in the different regions of the disk, for our fiducial model. Only a fraction of MBHB spend time in *Region-2* and *Region-3* disks, and even that is only for a small region of log-radius space. While almost all systems do enter *Region-1* by about  $10^3 R_s$ , GW hardening has, in general, also become significant by these same scales.

In addition to the spatially-distinct disk regions, different types of migration occur depending on whether the disk or the secondary-MBH is dynamically dominant (analogous to the distinction between



**Figure 1.8:** Hardening timescales due to circumbinary disk drag for binaries grouped by total mass and mass ratio. Light and heavy MBHB are separated by total masses below and above  $10^9 M_{\odot}$  respectively; and minor and major based on mass ratios ( $\mu$ ) below and above  $1/4$ . The median and surrounding 50% intervals for all MBHB systems are shown in grey, showing that ‘light’ systems dominate the bulk of the binary population. Heavy, and especially heavy-major, systems tend to harden orders of magnitude faster than lighter ones. The light population (especially major) exhibits nonmonotonicities at intermediate separations ( $\sim 10^{-3} - 10^{-1}$  pc) indicative of changes between disk regions. Heavy systems (especially minor), on the other hand, show much smoother hardening rates consistent with moving through primarily *Region-1*.

‘planet-dominated’ and ‘disk-dominated’, Type II migration in planetary disks—see, [HKM09](#)). If disk-dominated, the system hardens on the viscous timescale  $\tau_v$ , whereas if the secondary is dominant—as is the typical case in our simulations, the timescale is slowed by a factor related to the degree of secondary-dominance.

The resulting hardening timescales due to VD from a circumbinary disk are shown in Fig. 1.8. The more massive binaries (‘heavy’,  $M > 10^9 M_{\odot}$ ) have orders of magnitude shorter VD hardening times, but are quite rare. The overall trend (grey, cross-hatched) follows the less massive (‘light’) systems. The changes in slope of the ‘light’ populations (especially ‘Major’,  $\mu \equiv M_2/M_1 > 1/4$ ) at separations larger than  $10^{-3}$  pc are due to transitions in disk regions. The ‘heavy’ systems tend to have SG unstable Regions 2 & 3, and thus harden more smoothly, predominantly due to *Region-1*.

Fig. 1.9 compares median hardening rates in simulations including VD (solid) with those without a disk (dashed). In the former case, the purely VD hardening rates are also shown (dotted)—with the maximum disk cutoff  $\mathcal{R}_{\text{SG,Max}}$ , clearly apparent at 10 pc. Different dynamical friction prescriptions are shown, with mass enhancements over dynamical times calculated using the ‘stellar’ method in blue, and fixed 1 Gyr timescales in red (see §1.3.1). The upper panel shows a moderately refilled loss cone ( $\mathcal{F}_{\text{refill}} = 0.6$ ), while in the lower panel the LC is always full ( $\mathcal{F}_{\text{refill}} = 1.0$ ). The effects of VD are clearly apparent below a few  $10^{-2}$  pc in all models, and up to  $\mathcal{R}_{\text{SG,Max}} = 10$  pc when  $\mathcal{F}_{\text{refill}} = 0.6$ . In the  $\mathcal{F}_{\text{refill}} = 1.0$  case, LC hardening dominates to much smaller separations, making the VD effects minimal for the overall hard-

ening rates. This is echoed in the changes in coalescing fractions\* between the VD and No-VD cases: for  $\mathcal{F}_{\text{refill}} = 0.6$ , VD increases  $\mathcal{F}_{\text{coal}}$  by  $\sim 10\%$ , while for  $\mathcal{F}_{\text{refill}} = 1.0$ , it is only increased by  $\sim 2\%$ .

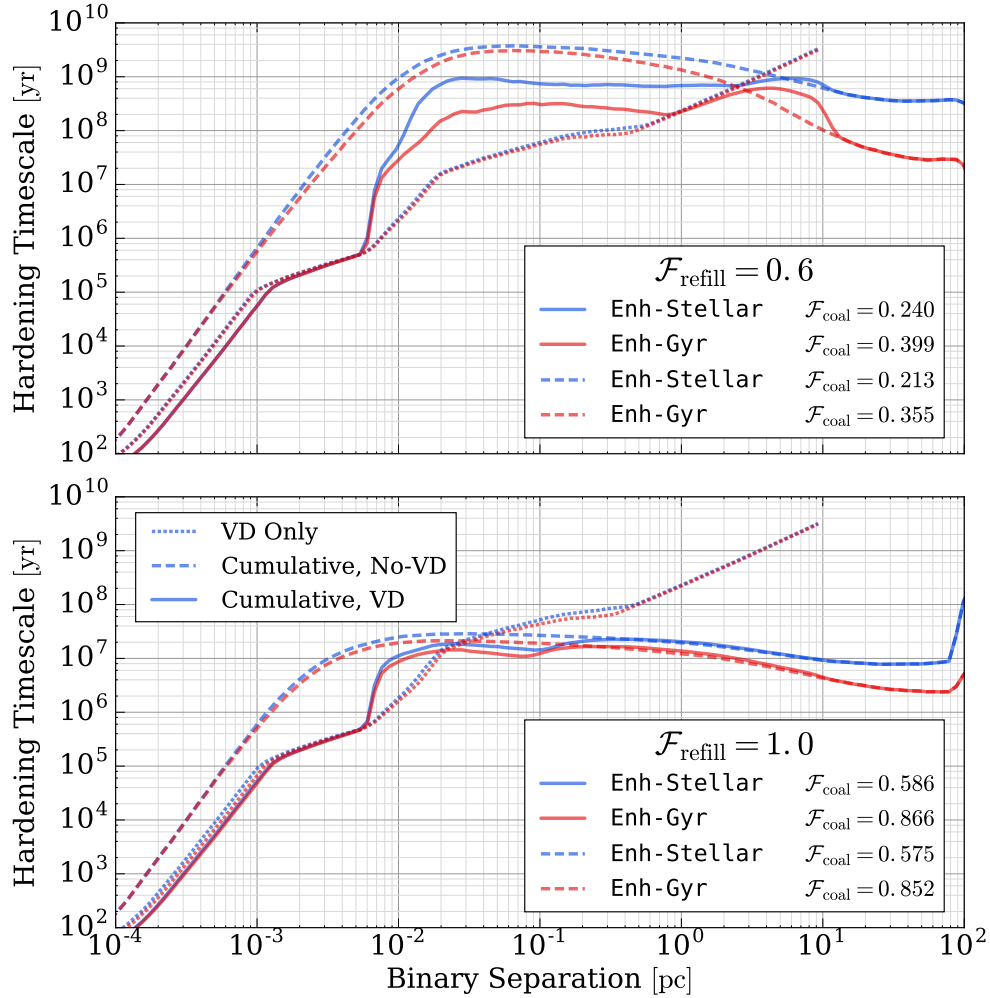
The circumbinary disk is SG unstable for many systems, and thus the median hardening rates including VD are often intermediate between the purely VD timescales and simulations without VD at all (e.g., seen between  $\sim 5 \times 10^{-3} - 1$  pc in the upper panel of Fig. 1.9). At smaller separations ( $\lesssim 10^{-3}$  pc), where LC is almost always subdominant to VD and/or GW hardening, the VD decreases the median hardening timescales by between a factor of a few, and an order of magnitude. Notably, Fig. 1.9 shows that the scaling of hardening rate with separation below about  $10^{-3}$  pc is very similar between that of GW (which is dominant in the No-VD case) and VD. At these small scales,  $\gtrsim 80\%$  of our binaries are in disk *Region-1* (see Fig. 1.7), which has a viscous hardening rate,  $\tau_{v,1} \propto r^{7/2}$  (HKM09, Eq. 21a), compared to a very similar scaling for GW,  $\tau_{\text{gw}} \propto r^4$  (see §1.3.4). Thus, even when VD dominates hardening into the mpc-scale regime, we don't expect the GWB spectrum from binaries in *Region-1* to deviate significantly from the canonical  $-2/3$  power-law.

Differences in median hardening timescales, solely due to viscous drag, are compared for a variety of VD parameters in Fig. 1.10. A simulation with our fiducial disk parameters is shown in dashed-black, and each color shows variations in a different parameter. Decreasing the viscosity of the disk ( $\alpha$ , green) amounts to a proportional increase in the hardening timescale, and decrease in the coalescing fractions ( $\mathcal{F}_{\text{coal}}$ ). Decreasing the maximum disk radius ( $\mathcal{R}_{\text{SG,Max}}$ , red) decreases the overall effectiveness of VD, but because gaseous DF continues in its place, the coalescing fraction remains unchanged. While  $\mathcal{R}_{\text{SG,Max}}$  changes the maximum disk radius,  $\lambda_{\text{sg}}$  changes the radii at which *Region-2* & *Region-3* become SG-unstable directly (i.e. even well within the maximum cutoff radius). Increasing  $\lambda_{\text{sg}}$  by a factor of four (blue) significantly increases the number of MBHB with SG-stable *Region-2*<sup>†</sup> between  $\sim 10^{-2}$  &  $10^{-1}$  pc, increasing the overall coalescing fraction.

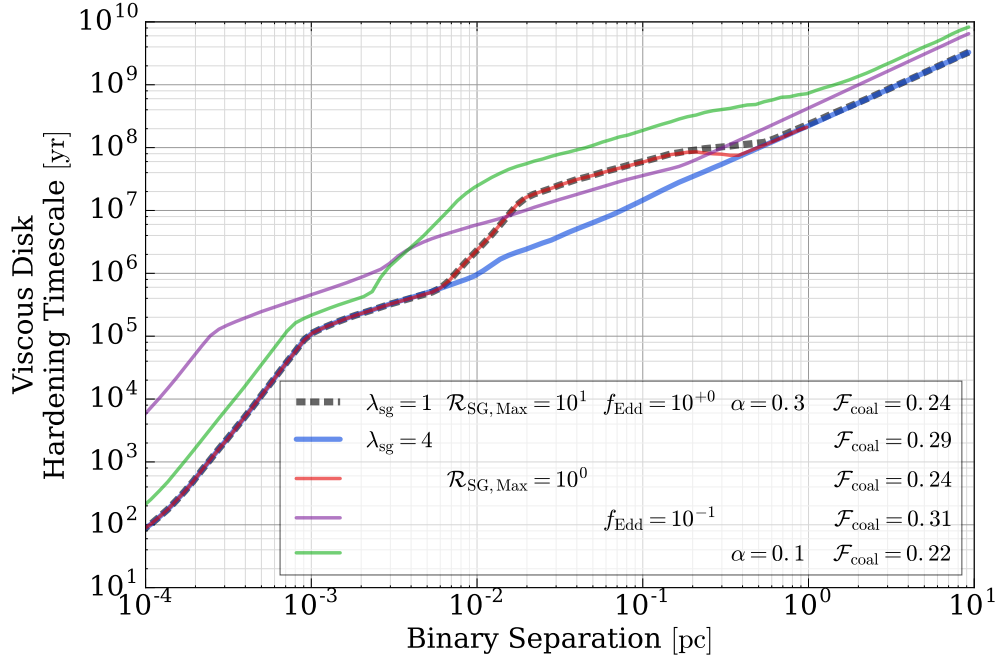
Decreasing the accretion rates ( $f_{\text{Edd}}$ , purple), and thus disk densities, increases the hardening timescales similarly to changing the viscosity (green). At the same time, significantly more systems have stable outer disks. This has the effect of increasing coalescing fractions noticeably, despite the increased median hardening timescales. In addition to increased outer-disk stability, the transition between disk regions are also inwards. A large number of MBHB at small separations ( $\lesssim 10^{-3}$  pc) remain in disk *Region-2* instead of transitioning to *Region-1*. This softens the scaling of hardening rate with separation to,  $\tau_{v,2} \propto r^{7/5}$  (HKM09, Eq. 21b), which differs much more significantly from purely GW-driven evolution.

\* $\mathcal{F}_{\text{coal}}$ , the fraction of systems with mass ratio  $\mu > 0.1$  which coalesce by  $z = 0$ .

<sup>†</sup>See the transition in Fig. 1.7 between *Region-2* (light-blue) & *Region-4* (grey) at  $\sim 10^{-2}$  pc.



**Figure 1.9:** Median Hardening timescales with and without drag from a circumbinary Viscous Disk (VD) for different hardening models. Different LC refilling parameters ( $\mathcal{F}_{\text{coal}} = 0.6$ , upper; and  $\mathcal{F}_{\text{coal}} = 1.0$ , lower) are compared against DF models (‘Enh-Stellar’, blue; ‘Enh-Gyr’, red). Each line type shows a different hardening rate: viscous drag only (dotted), and cumulative hardening rates with VD (solid) and without VD (dashed). With  $\mathcal{F}_{\text{refill}} = 0.6$ , VD effects are apparent up to the disk cutoff,  $\mathcal{R}_{\text{SG,Max}} = 10$  pc, whereas for  $\mathcal{F}_{\text{refill}} = 1.0$ , LC scattering dominates down to  $\sim 10^{-2}$  pc. Similarly, the presence of a circumbinary disk has a much more pronounced effect on the fraction of high mass-ratio ( $\mu > 0.1$ ) systems which coalesce by redshift zero ( $\mathcal{F}_{\text{coal}}$ ), which are indicated in the legends. At very low separations the cumulative (with-VD) hardening rate is very nearly the purely VD rate, showing its importance down to very small scales. At intermediate separations the ‘cumulative, VD’ rate is intermediate between the ‘cumulative, No-VD’ and the ‘VD only’ model, showing that the disk is only present in some fraction of systems at those scales.



**Figure 1.10:** Median Hardening timescales comparing our fiducial Viscous Disk (VD) parameters (black, dashed) with other configurations. The radius at which the disk becomes Self-Gravity (SG) unstable is  $\mathcal{R}_{\text{SG}} = \text{Min}[\lambda_{\text{sg}} r_Q, \mathcal{R}_{\text{SG,Max}}]$ , where  $r_Q$  is the radius at which the Toomre parameter reaches unity.  $\lambda_{\text{sg}}$  scales the SG-unstable radius, while  $\mathcal{R}_{\text{SG,Max}}$  is a maximum cutoff radius.  $\alpha$  is the standard disk viscosity parameter, and  $f_{\text{Edd}}$  limits the maximum accretion rate, i.e.  $\dot{M} = \text{Min}[\dot{M}_{\text{ill}}, f_{\text{Edd}}\dot{M}_{\text{Edd}}]$ . While this variety of VD parameters produces hardening rates varying by two orders of magnitude, the resulting changes to the coalescing fraction  $\mathcal{F}_{\text{coal}}$  is fairly moderate as VD is often subdominant to LC scattering at larger radii and to GW emission at smaller radii. Effects on  $\mathcal{F}_{\text{coal}}$  can be counterintuitive, for example decreasing the accretion rate (purple line) increases the median hardening timescale, but increases the coalescing fraction because the disk becomes SG-stable for a larger fraction of binaries. Each model uses  $\mathcal{F}_{\text{refill}} = 0.6$ , and the ‘Enh-Stellar’ DF.

### 1.3.4 GRAVITATIONAL-WAVE EMISSION

Gravitational wave radiation will always be the dominant dissipation mechanism at the smallest binary separations—within hundreds to thousands of Schwarzschild radii. GW hardening depends only on the constituent masses ( $M_1$  &  $M_2$ ) of the MBHB, their separation, and the system’s eccentricity. The hardening rate can be expressed as (Peters 1964),

$$\frac{da}{dt} = -\frac{64 G^3 M_1 M_2 (M_1 + M_2)}{5 c^5 a^3} \frac{(1 + \frac{73}{24} e^2 + \frac{37}{96} e^4)}{(1 - e^2)^{7/2}}, \quad (1.12)$$

where  $a$  is the semi-major axis of the binary and  $e$  is the eccentricity. In our treatment we assume that the eccentricities of all MBHB are uniformly zero, in which case, Eq. 1.12 can easily be integrated to find the time to merge,

$$t_{\text{GW}} = \frac{5c^5}{64G^3} \frac{a_0^4 - \mathcal{R}_{\text{crit}}^4}{M_1 M_2 M}, \quad (1.13)$$

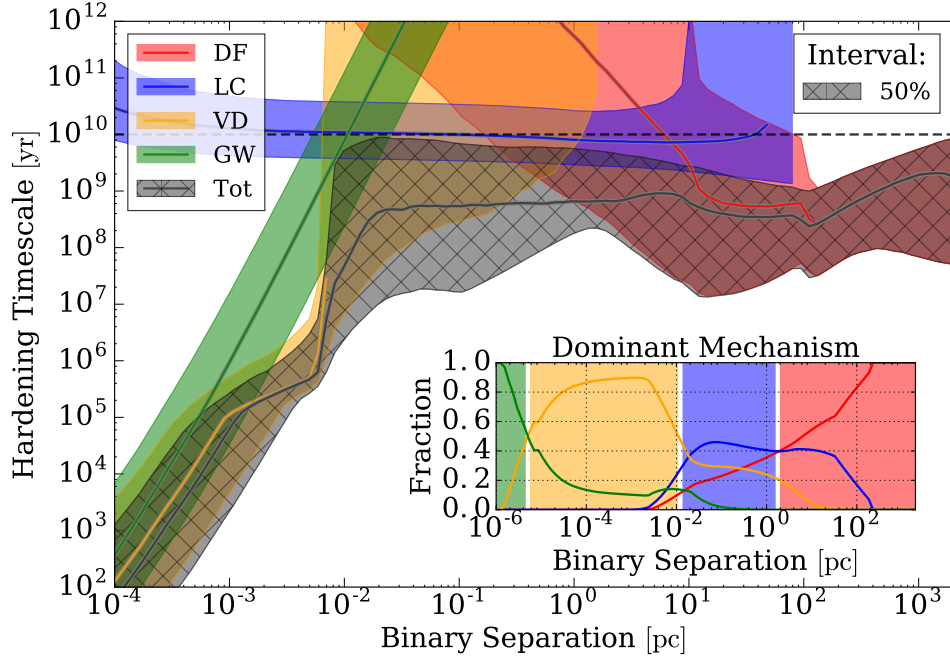
$$\approx 10^{10} \text{ yr} \left( \frac{a_0}{0.01 \text{ pc}} \right)^4 \left( \frac{M}{2 \times 10^7 M_\odot} \right)^{-3} \left( \frac{2 + \mu + 1/\mu}{4} \right),$$

for a total mass  $M = M_1 + M_2$ , mass ratio  $\mu \equiv M_2/M_1$ , initial separation  $a_0$  and critical separation  $\mathcal{R}_{\text{crit}}$ . In practice, we assume that the GW signal from binaries terminates at the Inner-most Stable Circular Orbit (ISCO), at which point the binary ‘coalesces’; i.e.  $\mathcal{R}_{\text{crit}} = \mathcal{R}_{\text{isco}}(J = 0.0) = 3R_s$ . For an equal mass binary, with median Illustris MBH masses\* of about  $10^7 M_\odot$ , the binary needs to come to a separation of  $\sim 0.01$  pc ( $\sim 10^4 R_s$ ), to merge within a Hubble time. Characteristic timescales and separations for (purely) GW-driven inspirals across total mass and mass ratio parameter space are plotted in Fig. 1 of Appendix A. While the absolute most-massive MBHB can merge purely from GW emission starting from a parsec, the bulk of physical systems, at  $10^6 - 10^8 M_\odot$ , must be driven by environmental effects to separations on the order of  $10^{-3} - 10^{-2}$  pc ( $\sim 500 - 5000 R_s$ ) to coalesce by redshift zero.

## 1.4 RESULTS

The hardening timescales for all Massive Black Hole Binaries (MBHB) are plotted against binary separation in Fig. 1.11, broken down by hardening mechanism. This is a representative model with a moderate loss-cone (LC) refilling fraction  $\mathcal{F}_{\text{refill}} = 0.6$  (see §1.3.2), using the ‘Enh-Stellar’ DF (see §1.3.1). This is the fiducial model for which we present most results, unless otherwise indicated. The inset shows the fraction of binaries with hardening rates dominated by each mechanism. DF is most important at large radii soon after binaries form, until LC scattering takes over at  $\sim 1$  pc. The median hardening timescale remains fairly consistent at a few times 100 Myr, down to  $\sim 10^{-2}$  pc at which point viscous drag (VD) drives the bulk of systems until gravitational wave (GW) emission takes over at separations below  $10^{-5}$  pc, where the typical hardening timescale reaches years. The landscape of hardening timescales for alternative DF prescriptions and LC refilling fractions are shown in the supplemental material (Fig. 1.24).

\*after typical selection cuts, described in §1.2.



**Figure 1.11:** Binary hardening timescales versus binary separation by mechanism. Colored lines and bands show the median and 50% intervals for Dynamical Friction (DF), Loss Cone (LC) scattering, Viscous Drag (VD), and Gravitational Wave (GW) emission with the total hardening rate shown by the grey, hatched region. The inset panel shows the fraction of binaries dominated by each mechanism. This simulation uses our fiducial parameters (e.g.  $\mathcal{F}_{\text{refill}} = 0.6$ ), with ‘Stellar’ DF-mass enhancement. The binary hardening landscape is very similar to that outlined by BBR80, but the details are far more nuanced. For a comparison with alternate models, see Fig. 1.24.

#### 1.4.1 BINARY LIFETIMES

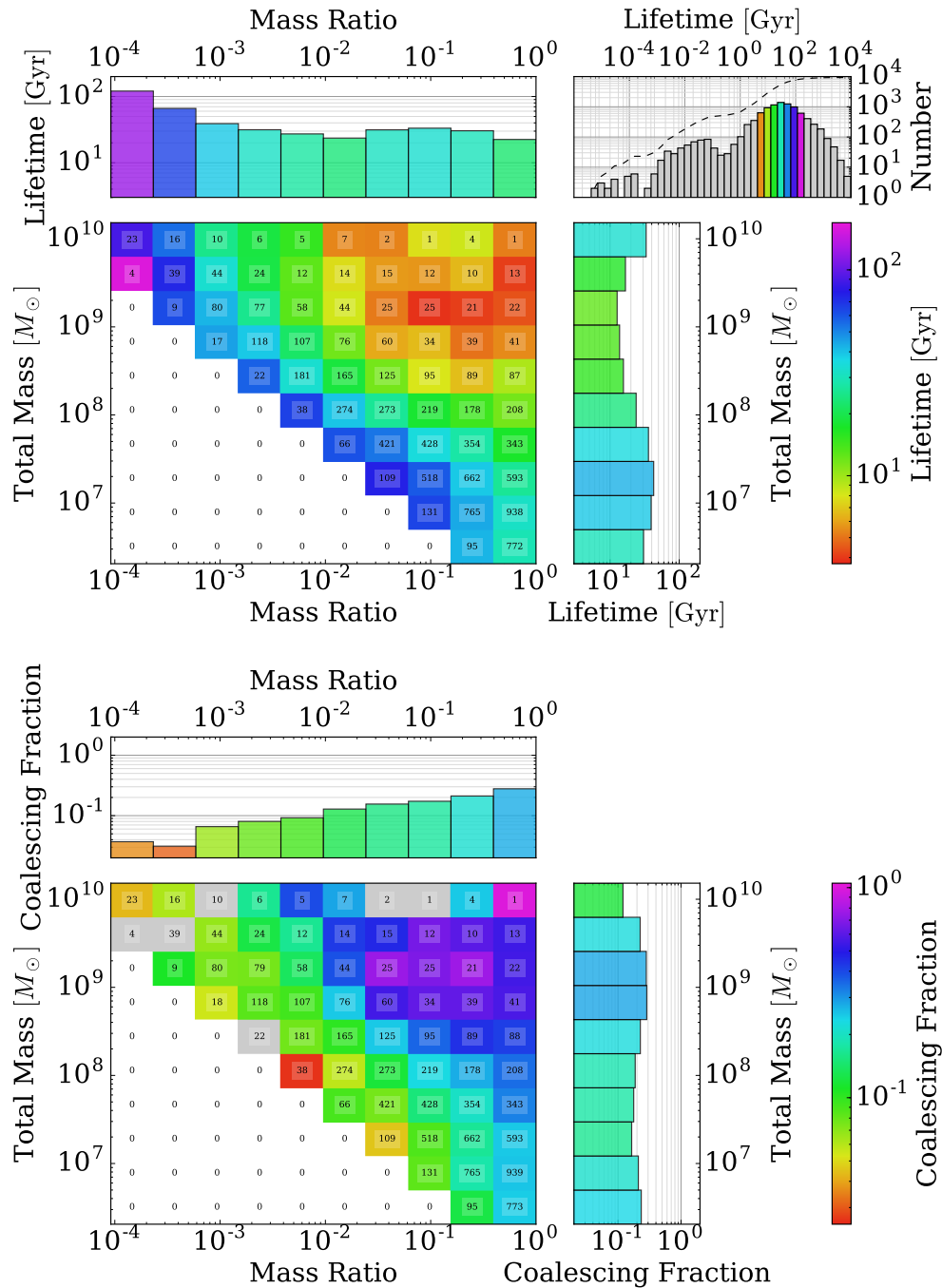
Characteristic hardening timescales are often many 100 Myr, and MBHB typically need to cross eight or nine orders of magnitude of separation before coalescing. The resulting lifetimes of MBHB can thus easily reach a Hubble time. Fig. 1.12 shows binary lifetimes (upper panels) and the fraction of systems which coalesce by  $z = 0$  (lower panels) for our fiducial model. Systems are binned by total mass and mass ratio, with the number of systems in each bin indicated. The plotted lifetimes are median values for each bin, with the overall distribution shown in the upper-right-most panel. Grey values are outside of the range of binned medians, and the cumulative distribution is given by the dashed line.

The lifetime distribution peaks near the median value of 29 Gyr, with only  $\sim 7\%$  of lifetimes at less than 1 Gyr. About 20% of all MBHB in our sample coalesce before redshift zero. Systems involving the lowest mass black holes\* (i.e. down and left) tend towards much longer lifetimes. Overall, lifetimes and coalescing fractions are only mildly correlated with total mass or mass ratio, when marginalizing over the other. For systems with total masses  $M > 10^8 M_{\odot}$ , the coalescing fraction increases to 23%, and for mass ratios  $\mu > 0.2$ , only slightly higher to 26%. In general, examining slightly different total-mass or mass-ratio cutoffs has only minor effects on lifetimes and coalescing fractions.

There is a strong trend towards shorter lifetimes for simultaneously high total masses and mass ratios (i.e. up and right), where median lifetimes are *only* a few gigayear. Considering both  $\mu > 0.2$  and at the same time  $M > 10^8 M_{\odot}$ , coalescing fractions reach 45%. The handful of MBHB which coalesce after

\*Recall we require MBH masses of at least  $10^6 M_{\odot}$





**Figure 1.12:** Binary lifetimes (upper) and coalescing fractions (lower) for our fiducial model with a moderate DF and LC refilling (‘Enh-Stellar’ and  $\mathcal{F}_{\text{refill}} = 0.6$ , respectively). The overall distribution of MBHB lifetimes are shown in the upper-right-most panel, with the cumulative distribution plotted as the dashed line. The median lifetime is  $\sim 30$  Gyr overall, but is significantly shorter for MBHB with either high total masses, or nearly-equal mass ratios. For this group, the coalescing fractions are near unity. Grey bins in the lower panel correspond to those with no binaries which coalesce by redshift zero. While systems with the highest masses and mass ratios tend to have much shorter lifetimes, they also form at low redshifts with less time to coalesce.

$\lesssim 1$  Myr ( $\sim 0.3\%$ ) tend to involve MBH in over-massive galaxies (i.e. galaxy masses larger than expected from MBH-host scalings) with especially concentrated stellar and/or gas distributions. There are a handful of high mass ratio, and highest total mass MBHB ( $M \sim 10^{10} M_{\odot}$ ) systems showing a noticeable decrease in coalescing fraction. These systems form at low redshifts and don't have time to coalesce despite relatively short lifetimes.

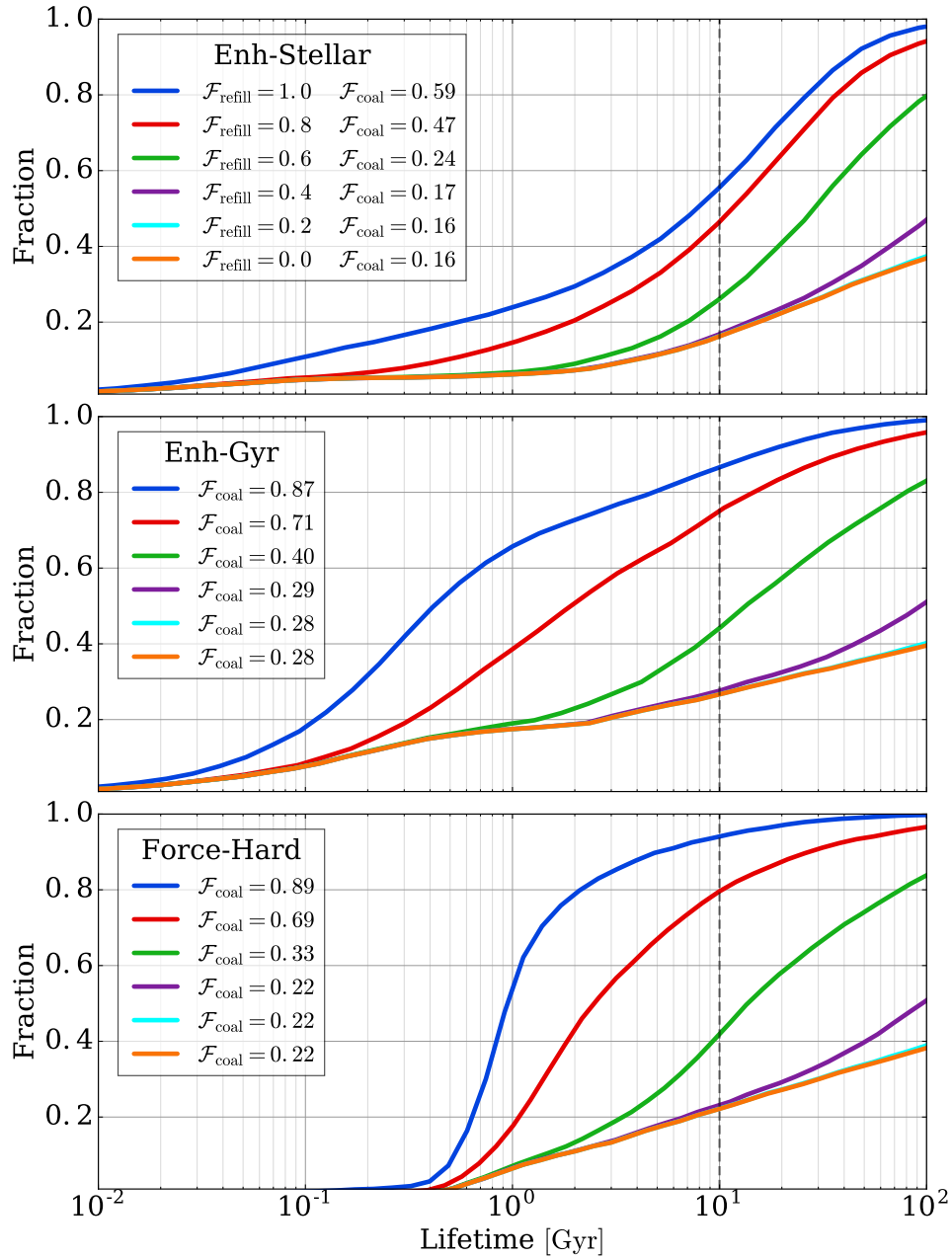
Lifetimes and coalescing fractions for an always full LC ( $\mathcal{F}_{\text{refill}} = 1.0$ ) are shown in Fig. 1.25, in Appendix A. The median value of the lifetime distribution shifts down to  $\sim 8$  Gyr, with  $\sim 24\%$  under 1 Gyr. The coalescing fractions increase similarly, and systems which are *either* high mass ratio ( $\mu \gtrsim 0.2$ ) *or* high total mass ( $M \gtrsim 10^8 M_{\odot}$ ) generally coalesce by redshift zero. Specifically, the coalescing fractions are 50% and 61% for all systems and those with  $\mu > 0.2$  respectively. Considering only  $M > 10^8 M_{\odot}$ , fractions increase to 54% & 99%.

Cumulative distributions of MBHB lifetimes are compared in Fig. 1.13 for a variety of LC refilling factors (colors) and our three primary DF prescriptions (panels; see §1.3.1). The first two panels correspond to prescriptions where the effective masses used in the DF calculation are the sum of the secondary MBH mass and the mass of its host galaxy. To model stripping of the secondary galaxy during the merger process, the effective mass decreases as a power law to the 'bare' MBH mass after a dynamical time. The 'Enh-Stellar' model (upper) calculates the dynamical time at twice the stellar half-mass radius, and the mass there enclosed. The 'Enh-Gyr' model (middle), on the other hand, uses a fixed 1Gyr timescale—almost a factor of ten longer than the median 'stellar'-calculated value. Finally, the 'Force-Hard' model (lower), uses the 'bare' secondary MBH mass, but the binary is forced to the hard binary regime (generally 1 – 10 pc) over the course of a dynamical time (calculated in the 'stellar' manner). Each color of line indicates a different LC refilling fraction, from always full ( $\mathcal{F}_{\text{refill}} = 1.0$ ; blue) to the steady state ( $\mathcal{F}_{\text{refill}} = 0.0$ ; orange). The fraction of high mass-ratio ( $\mu > 0.1$ ) systems which coalesce by redshift zero ( $\mathcal{F}_{\text{coal}}$ ) are also indicated in the legends.

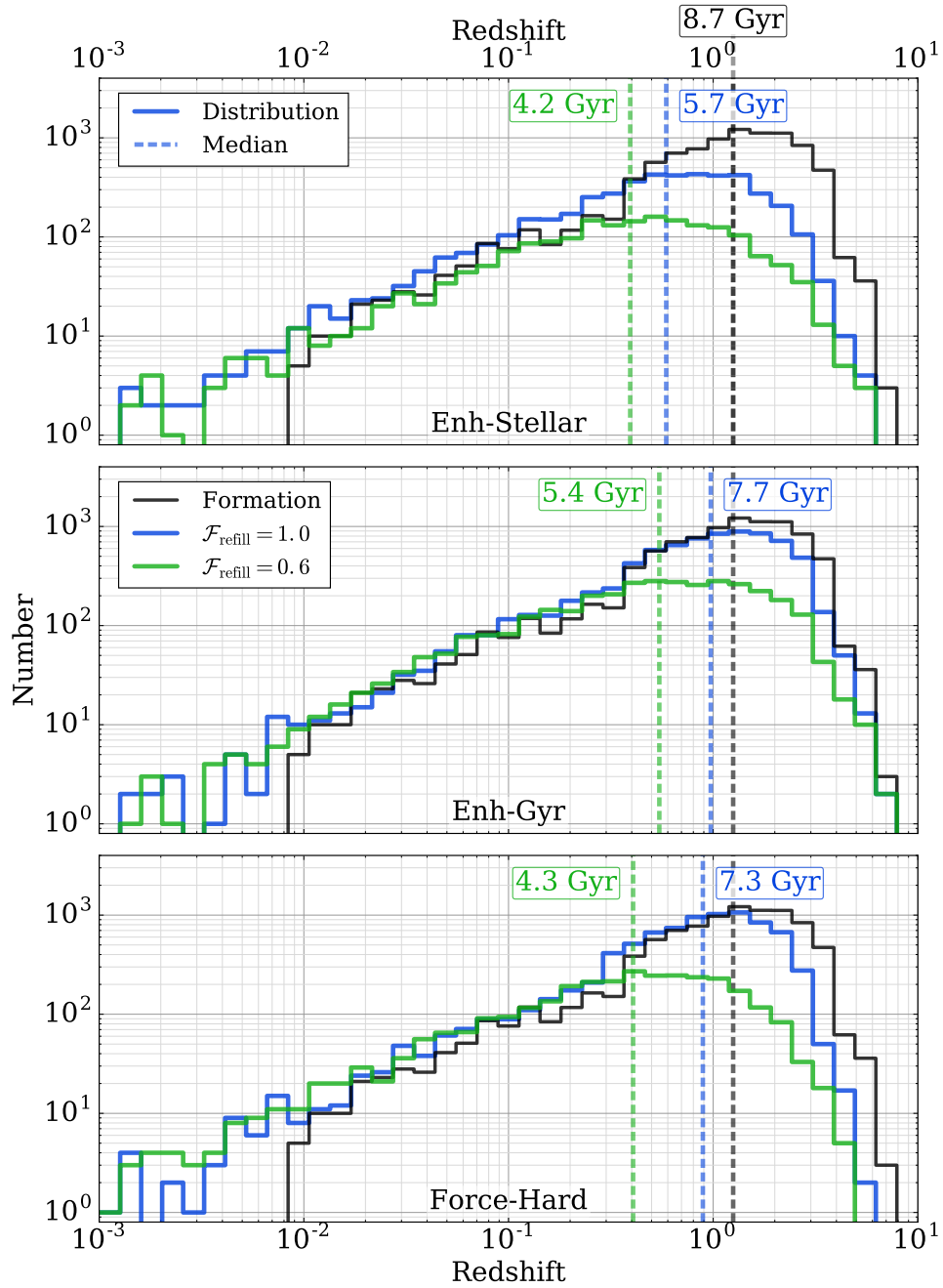
The high mass-ratio coalescing fractions tend to vary by almost a factor of four depending on the LC state, while the varying DF prescriptions have less than factor of two effect. Median lifetimes change considerably, however, even between DF models, for example with  $\mathcal{F}_{\text{refill}} = 1.0$ , the median lifetime for the 'Enh-Stellar' model is 7.7 Gyr, while that of 'Enh-Gyr' is only about 0.42 Gyr. Apparently, with a full LC, DF at large scales tends to be the limiting factor for most systems. While the highest overall  $\mathcal{F}_{\text{coal}}$  occurs for 'Force-Hard' &  $\mathcal{F}_{\text{refill}} = 1.0$ , it takes almost an order of magnitude longer for the first  $\sim 10\%$  of systems to coalesce than in either of the 'Enh' models. The effects of DF on the lifetimes of the first systems to merge are fairly insensitive to the LC state. There are thus cases where DF can be effective at driving some systems to coalesce very rapidly. At the same time, for the bulk of systems, after hardening past kiloparsec scales the remaining lifetime can be quite substantial. For  $\mathcal{F}_{\text{refill}} \lesssim 0.6$ , neither the precise LC refilling fraction nor the DF model make much of a difference after the first 10 – 30% of systems coalesce. In these cases, the most massive systems with high-mass ratios coalesce fairly rapidly regardless, but the smaller more extreme mass-ratio systems take many Hubble times to merge.

Figure 1.14 shows the distribution of formation (black) and coalescence (colored) redshifts resulting from a variety of binary evolution models. Each panel shows a different DF prescription, and two LC refilling parameters are shown: always full,  $\mathcal{F}_{\text{refill}} = 1.0$  (blue), and our fiducial, moderately refilled value of  $\mathcal{F}_{\text{refill}} = 0.6$  (green). A handful of events have been cutoff at low redshifts ( $z < 10^{-3}$ ) where the finite volume of the Illustris simulations and cosmic variance becomes important. Median redshifts for each distribution are overplotted (dashed), along with their corresponding look-back times. The median formation redshift for our MBHB is  $z = 1.25$  (look-back time of  $\sim 8.7$  Gyr). For a full LC and the stronger DF models, 'Enh-Gyr' and 'Force-Hard', the median coalescence redshifts are delayed to  $z \sim 1.0$  and  $z \sim 0.9$

respectively—i.e. by about a gigayear. For our more modest, fiducial DF prescription, ‘Enh-Stellar’, even the full LC case still delays the median coalescence redshift to  $z \sim 0.6$ , about 3 Gyr after the peak of MBHB formations. If the LC is only moderately refilled, the median redshifts are much lower: between  $z \sim 0.4 - 0.6$ .



**Figure 1.13:** Cumulative distributions of binary lifetimes for a variety of DF and LC parameters. The ‘Enh-Stellar’, ‘Enh-Gyr’, and ‘Force-Hard’ DF models are shown in each panel, and the colors of lines indicate the LC refilling fraction. The fraction of high mass ratio ( $\mu > 0.1$ ) systems coalescing by  $z = 0$  are given in the legend ( $\mathcal{F}_{\text{coal}}$ ).  $\mathcal{F}_{\text{refill}}$  is the dominant factor determining the lifetime distribution, but the DF model significantly affects the earliest merging systems, and overall fraction of coalescing systems.



**Figure 1.14:** Distribution of MBH binary formation (black) and coalescence (colored) redshifts for different DF models (panels) and two LC refilling parameters: always full ( $\mathcal{F}_{\text{refill}} = 1.0$ ; blue) and our fiducial, moderately refilled ( $\mathcal{F}_{\text{refill}} = 0.6$ ; green) value. The median redshift for each distribution is also plotted (dashed), with the corresponding look-back time indicated. The minimum delay time between medians of formation and coalescence is 1 Gyr, but up to 4.5 Gyr for our fiducial LC state and DF model (‘Enh-Stellar’).

### 1.4.2 THE GRAVITATIONAL WAVE BACKGROUND (GWB)

In §1.1 we have outlined the theoretical background for the existence of a stochastic GWB, and introduced the formalism for calculating pure power-law spectra. Fig. 1.15 shows the purely power-law spectrum derived from Illustris MBH binaries, assuming that all systems (passing our selection cuts outlined in §1.2) reach the GW dominated regime and evolve purely due to GW emission. Other representative power-law predictions (see §1.1) and recent pulsar timing array (PTA) upper limits are included for comparison. The Illustris prediction is completely consistent with the existing literature and about 30% below the most recent PTA upper limits. These consistencies validate the Illustris MBHB population, and the prescriptions for the growth and evolution of individual MBH.

Almost all of the details of binary evolution are obscured in purely power-law predictions (i.e. Eq. 1.1). In particular, they imply that all MBHB instantly reach the separations corresponding to the frequencies of interest, and evolve purely due to GW-emission. In reality, we've shown that the delay time distribution can be significant at fractions of a Hubble time. This has the important consequence that not all MBHB coalesce (or even reach the PTA band) before redshift zero. At the same time, the environmental effects (e.g. LC scattering) which are *required* to bring MBH binaries to the relevant orbital frequencies also decrease the time they emit in each band, attenuating the GW signal.

#### 1.4.2.1 FULL GWB CALCULATION FORMALISM

The GWB can be calculated more explicitly by decomposing the expression for GW energy radiated per logarithmic frequency interval,

$$\frac{d\varepsilon_{\text{GW}}}{d \ln f_r} = \frac{d\varepsilon_{\text{GW}}}{dt_r} \frac{dt_r}{d \ln f_r}, \quad (1.14)$$

where the right-hand-side terms are the GW power radiated and the time spent in each frequency band.

The latter term can be further rewritten using Kepler's law as,

$$\frac{dt_r}{d \ln f_r} = f_r \left( \frac{df_r}{dt_r} \right)^{-1} = \frac{3}{2} \frac{a}{da/dt_r}, \quad (1.15)$$

where 'a' is the semi-major axis of the binary. In this expression, we can identify the binary 'hardening time',  $\tau_h \equiv a / (da/dt_r)$ . For reference, the binary separations corresponding to each GW frequency are shown in 2. While the GW power radiated is determined solely by the binary configuration (chirp mass and orbital frequency), the hardening time is determined by both GW emission and the sum of all environmental hardening effects. For more generalized binary evolution we can write,

$$\frac{d\varepsilon_{\text{GW}}}{d \ln f_r} = \frac{d\varepsilon_{\text{GW}}}{d \ln f_r} \Big|_{\text{GW}} \frac{\tau_h}{\tau_{\text{gw}}}. \quad (1.16)$$

This can be used to reformulate the GWB spectrum calculation<sup>†</sup> (Eq. 1.1) as,

$$h_c^2(f) = \frac{4\pi}{3c^2} (2\pi f)^{-4/3} \int \frac{(GM)^{5/3}}{(1+z)^{1/3}} \frac{\tau_h}{\tau_{\text{gw}}} \frac{d^3n}{dz d\mathcal{M} d\mu} dz d\mathcal{M} d\mu, \quad (1.17)$$

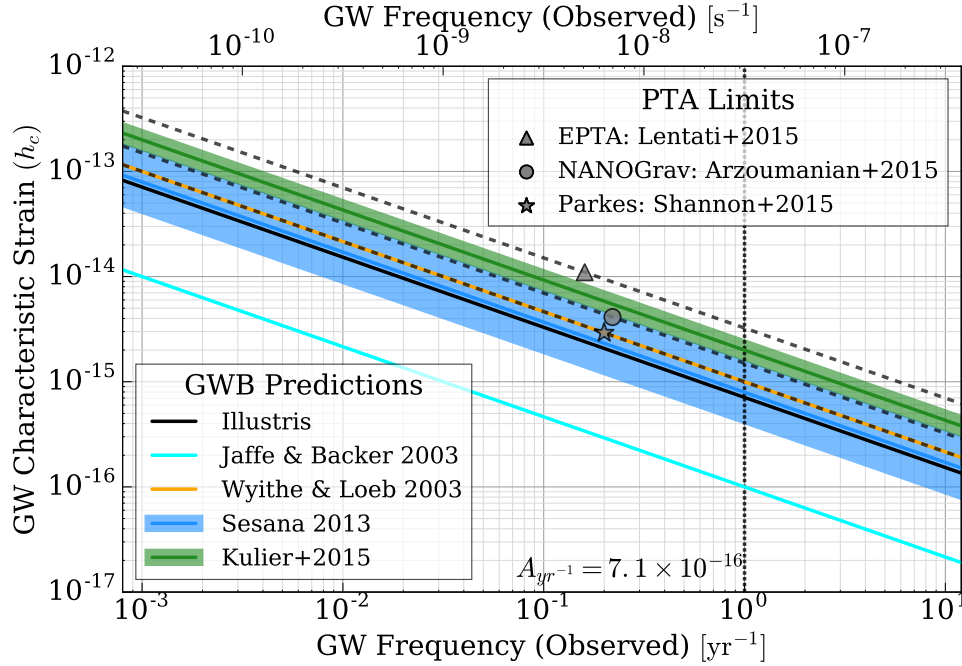
or for discrete sources,

$$h_c^2(f) = \frac{4\pi}{3c^2} (2\pi f)^{-4/3} \sum_i \frac{(GM_i)^{5/3}}{V_c (1+z_i)^{1/3}} \frac{\tau_{r,i}}{\tau_{\text{gw},i}}. \quad (1.18)$$

Additional hardening mechanisms will decrease the hardening timescale, i.e.  $\tau_h/\tau_{\text{gw}} \leq 1$ , decreasing the GWB. The purely power-law expression in Eq. 1.1 (and the Illustris spectrum in Fig. 1.15) thus represents an upper-limit to the GWB amplitude. While non-GW mechanisms are required to bring MBH binaries

<sup>\*</sup>Sometimes called the 'residence time' in the context of GW spectra.

<sup>†</sup>For a more complete derivation, see Kocsis & Sesana (2011).



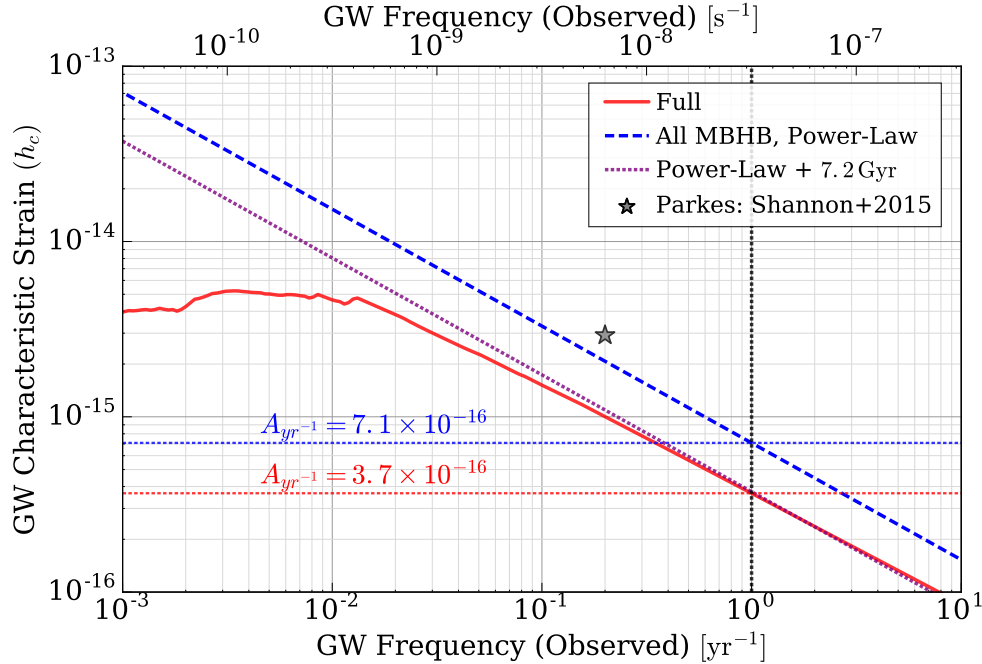
**Figure 1.15:** Stochastic Gravitational Wave Background spectrum produced by Illustris MBH binaries, assuming purely power-law evolution with all systems efficiently reaching the GW regime. Power-law predictions from the literature (described in §1.1) are presented for comparison, along with the most recent PTA upper limits. The power-law spectrum resulting from the Illustris simulations is very consistent with previous results, and about 30% below the most stringent observational upper limits.

close enough to effectively emit gravitational waves, they also attenuate the amplitude of the GW background.

#### 1.4.2.2 FIDUCIAL MODEL PREDICTIONS

The stochastic Gravitational Wave Background (GWB) resulting from our fiducial model is presented in Fig. 1.16. The ‘Full’ calculation (red, solid) uses Eq. 1.18, including the effects of DF, LC scattering, and VD in addition to GW emission. This is compared to a purely power-law model (blue, dashed), calculated with Eq. 1.1 and assuming that all Illustris MBHB reach the PTA-band rapidly, and evolve solely due to GW-emission. The amplitudes at  $1 \text{ yr}^{-1}$  are indicated, showing that the full hardening calculation with an amplitude of  $A_{\text{yr}^{-1}} \approx 3.7 \times 10^{-16}$  amounts to an almost 50% decrease from the naive, power-law estimate of  $A_{\text{yr}^{-1}} \approx 7.1 \times 10^{-16}$ .

The amplitude of the full GWB calculation can be matched at  $1 \text{ yr}^{-1}$  using the power-law model by introducing a uniform delay time of  $\sim 7.2 \text{ Gyr}$ —such that the systems which formed within a look-back time of 7.2 Gyr don’t coalesce or reach the relevant frequency ranges. This is shown in Fig. 1.16 (purple, dotted) as a heuristic comparison. At frequencies of the PTA band ( $\sim 0.1 \text{ yr}^{-1}$ ) and higher, our full calculation very nearly matches the  $A_{\text{yr}^{-1}} \propto f^{-2/3}$  power-law. A significant flattening of the spectrum is apparent at and below a few  $10^{-2} \text{ yr}^{-1}$ , where environmental effects (e.g. LC-scattering) significantly increase the rate at which MBHB move through a given frequency band, decreasing  $\tau_{\text{h}}$  and thus attenuating the amplitude of the GWB. The particular location and strength of the spectral flattening (or turnover) depends on the details of the DF and especially LC models.



**Figure 1.16:** Stochastic Gravitational Wave Background calculated from Illustris MBH binaries. The ‘Full’ calculation, shown in red, includes environmental effects from dynamical friction (‘Enh-Stellar’), stellar scattering ( $\mathcal{F}_{\text{refill}} = 0.6$ ), and a viscous circumbinary disk. Purely power-law models are also shown, for all Illustris MBHB (blue, dashed) and only the MBHB which coalesce by redshift zero after being delayed for 7.2 Gyr (purple, dotted). The GWB strain amplitudes at the standard frequency of  $1 \text{ yr}^{-1}$  are given, showing that a complete model of MBHB evolution leads to a  $\sim 50\%$  decrease of the signal. The most stringent observational upper limits are also shown.

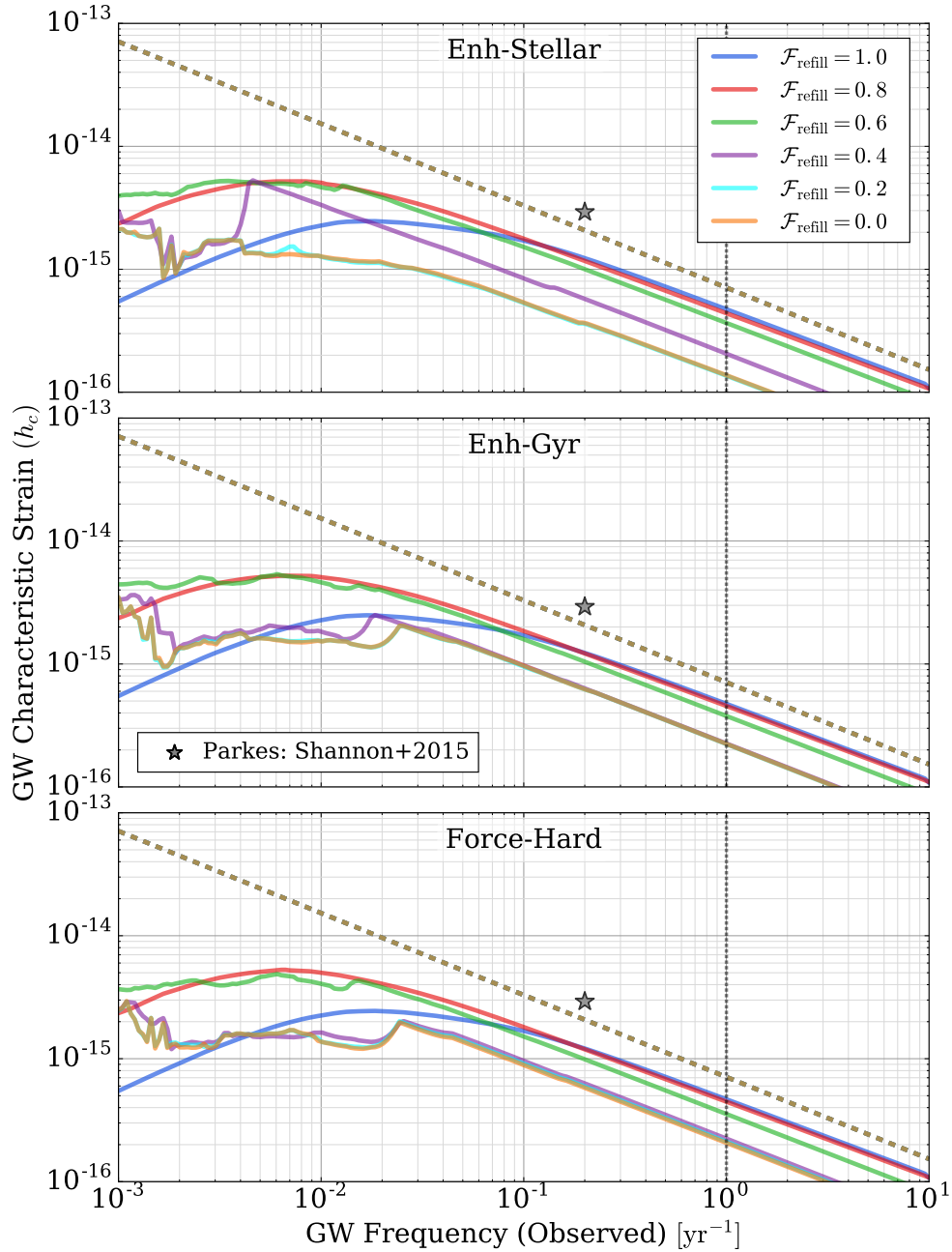
#### 1.4.2.3 GWB VARIATIONS WITH DYNAMICAL FRICTION AND LOSS-CONE MODEL PARAMETERS

Figure 1.17 compares the GWB spectrum for different DF prescriptions (panels) and LC refilling fractions (line colors). The naive, power-law model is shown as the dashed line for comparison, along with the most stringent PTA upper limit. Effects from variations in the DF prescription are strongly subdominant to changes in the LC state. The spectral shape is determined almost entirely, and at times sensitively, to  $\mathcal{F}_{\text{refill}}$ . For  $\mathcal{F}_{\text{refill}} < 0.8$ , the spectrum flattens at low frequencies, whereas for higher values it becomes a turnover. Even then, the location of the peak amplitude of the spectrum changes by more than a factor of two between  $\mathcal{F}_{\text{refill}} = 0.8$  and  $\mathcal{F}_{\text{refill}} = 1.0$ .

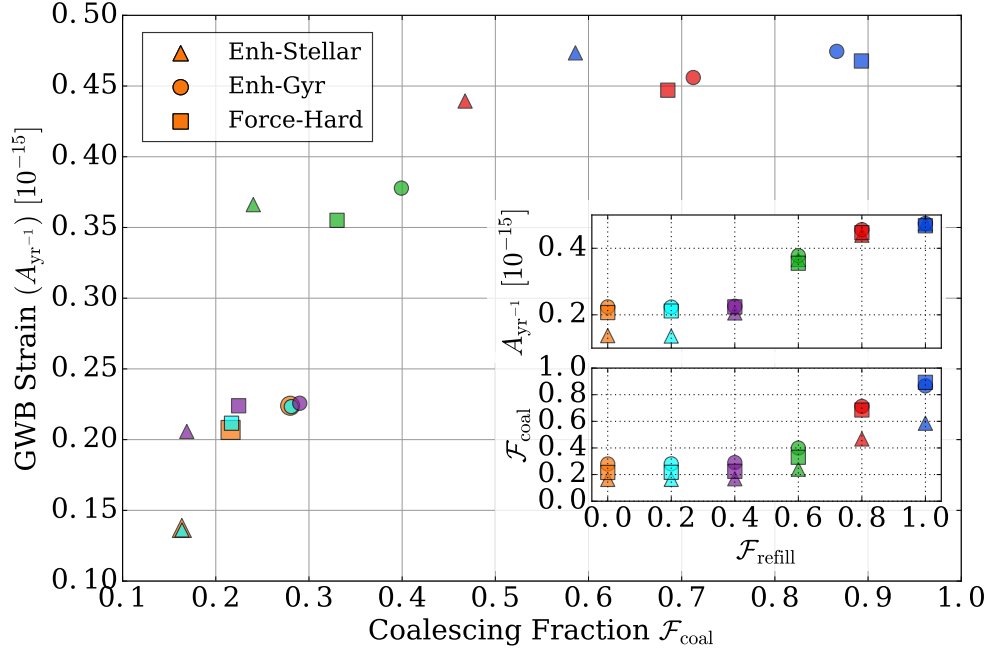
The cutoff seen in the full LC case ( $\mathcal{F}_{\text{refill}} = 1.0$ ) is very similar to that found by [Sesana \(2013a\)](#) (with ours  $\sim 5$  times lower amplitude), who show that in the scattering-dominated regime the GWB turns into a  $h_c \propto f$  spectrum. [McWilliams, Ostriker & Pretorius \(2014\)](#) also find a spectral cutoff, but at an order of magnitude higher frequency and amplitude. Unlike the results of [Ravi et al. \(2014\)](#), the cutoffs in our predicted GWB spectra are always at lower frequencies than will be reached by PTA in the next decade or so, likely because we assume zero eccentricity in the binary evolution. In the near future we hope to present results expanded to include eccentric evolution, in addition to exploring ‘deterministic’ or ‘continuous’ GW sources—i.e. sources individually resolvable by future PTA observations.

For each DF case in Fig. 1.17, the GWB spectrum is almost identical between  $\mathcal{F}_{\text{refill}} = 0.0, 0.2$  &  $0.4$ , with very little change in the coalescing fractions. This is consistent with changes in the distribution of lifetimes from varying DF and LC parameters. Looking at  $f = 1 \text{ yr}^{-1}$ , there is a sudden jump in am-





**Figure 1.17:** Comparison of GWB spectrum with variations in the LC refilling parameter and DF model. Each panel shows a different DF model, and each line-color a different  $\mathcal{F}_{\text{refill}}$ . Solid lines show the full GWB calculation while the dashed lines show the power-law model using all Illustris MBHB. Variations in  $\mathcal{F}_{\text{refill}}$  have much stronger effects on the spectrum than the DF model, changing the location and strength of the spectral break at lower frequencies. There tends to be a substantial jump in GWB amplitude between  $\mathcal{F}_{\text{refill}} = 0.4$  and  $0.6$ , with more gradual variations on either side. The full range of amplitudes at  $1 \text{ yr}^{-1}$  and  $10^{-1} \text{ yr}^{-1}$  are  $A_{\text{yr}^{-1}} = 0.14 - 0.47 \times 10^{-15}$  and  $A_{0.1\text{yr}^{-1}} = 5.4 - 17 \times 10^{-15}$ .



**Figure 1.18:** Dependence of GWB strain amplitude and binary coalescing fraction on LC refilling parameter and DF model. The GW strain is measured at the canonical  $f = 1 \text{ yr}^{-1}$ , and the coalescing fraction is defined using the population of high mass-ratio  $\mu > 0.1$  systems. Each symbol represents a different DF model, and each color a different LC refilling parameter.  $A_{\text{yr}^{-1}}$  tends to increase monotonically with  $\mathcal{F}_{\text{coal}}$ , but plateaus above  $\mathcal{F}_{\text{coal}} \gtrsim 0.5$ . The insets show how each  $A_{\text{yr}^{-1}}$  and  $\mathcal{F}_{\text{coal}}$  change with  $\mathcal{F}_{\text{refill}}$ . The strongest changes in  $\mathcal{F}_{\text{refill}}$  and  $A_{\text{yr}^{-1}}$  occur at slightly different values of the refilling fraction. This is because for an increase in  $\mathcal{F}_{\text{refill}}$ , the additional MBHB which are then able to coalesce tend to be the most massive of those which were previously persisting. At  $\mathcal{F}_{\text{refill}} \approx 0.5$  there is a significant change  $A_{\text{yr}^{-1}}$ , due to more massive and stronger GW emitting MBHB merger at that point. At  $\mathcal{F}_{\text{refill}} \approx 0.7$ ,  $\mathcal{F}_{\text{coal}}$  changes significantly due to less massive MBHB then being able to merger, and them constituting a larger portion of the binary population.

plitude with  $\mathcal{F}_{\text{refill}} = 0.6$ , and a modest increase in the coalescing fraction. Between  $\mathcal{F}_{\text{refill}} = 0.6$  and  $\mathcal{F}_{\text{refill}} = 0.8$ , on the other hand, there tends to be a more modest increase in GWB amplitude, but a roughly factor of two increase in  $\mathcal{F}_{\text{coal}}$ . This contrast arises from the changing population of MBHB which are brought to coalescence from each marginal change in refilling fraction. For an increase in  $\mathcal{F}_{\text{refill}}$ , the additional MBHB which are then able to coalesce tend to be the most massive of those which were previously persisting. Those, more massive systems, then have a larger effect on the GWB

Figure 1.18 shows the strain at  $1 \text{ yr}^{-1}$  ( $A_{\text{yr}^{-1}}$ ) versus coalescing fraction for the same set of DF and LC models. The colors again show different  $\mathcal{F}_{\text{refill}}$ , and now symbols are used for different DF prescriptions. The GWB amplitude is strongly correlated with coalescing fraction, but plateaus once roughly 50% of high mass-ratio MBHB are coalescing. Different DF parameters have little effect on  $A_{\text{yr}^{-1}}$  but more noticeably affect  $\mathcal{F}_{\text{coal}}$ , in both cases this is especially true at higher  $\mathcal{F}_{\text{refill}}$ . The inset panels show, independently, how  $A_{\text{yr}^{-1}}$  and  $\mathcal{F}_{\text{coal}}$  scale with  $\mathcal{F}_{\text{refill}}$  and DF model, reinforcing the previous points. In general, as  $\mathcal{F}_{\text{refill}}$  increases, lower total-mass MBHB systems are able to reach the PTA band, contribute to the GWB and coalesce effectively. At  $\mathcal{F}_{\text{refill}} \approx 0.5$ , the large increase in  $A_{\text{yr}^{-1}}$  is driven by massive MBH coming to coalescence, while at  $\mathcal{F}_{\text{refill}} \approx 0.7$ , a large number of MBH at lower masses are driven together, significantly increasing the coalescing fraction, but only marginally increasing  $A_{\text{yr}^{-1}}$ .

At the higher frequencies just discussed the GWB strain increases monotonically with  $\mathcal{F}_{\text{refill}}$  and coalescing fraction. This is intuitive as increasing effectiveness of the LC means more MBHB are able to reach the GW-regime and then coalesce. Fig. 1.17 shows that this trend is not the case at lower frequencies (i.e.  $f \lesssim 10^{-1} \text{ yr}^{-1}$ )—where the highest  $\mathcal{F}_{\text{refill}}$  show a decrease in the GWB amplitude. This can be seen more clearly in Fig. 1.26, which shows the GWB amplitude at  $f = 10^{-2} \text{ yr}^{-1}$  versus coalescing fraction. The trend is generally the same—strain increasing with  $\mathcal{F}_{\text{refill}}$ —until  $\mathcal{F}_{\text{refill}} = 1.0$  at which point the GWB amplitude drops significantly. At these low frequencies, LC stellar scattering is effective enough to significantly attenuate the GWB amplitude. This reflects a fundamental tradeoff in the realization of environmental effects: on one side bringing more MBHB into a given frequency bands, at the same time as driving their evolution rapidly through it, and attenuating the GW signal.

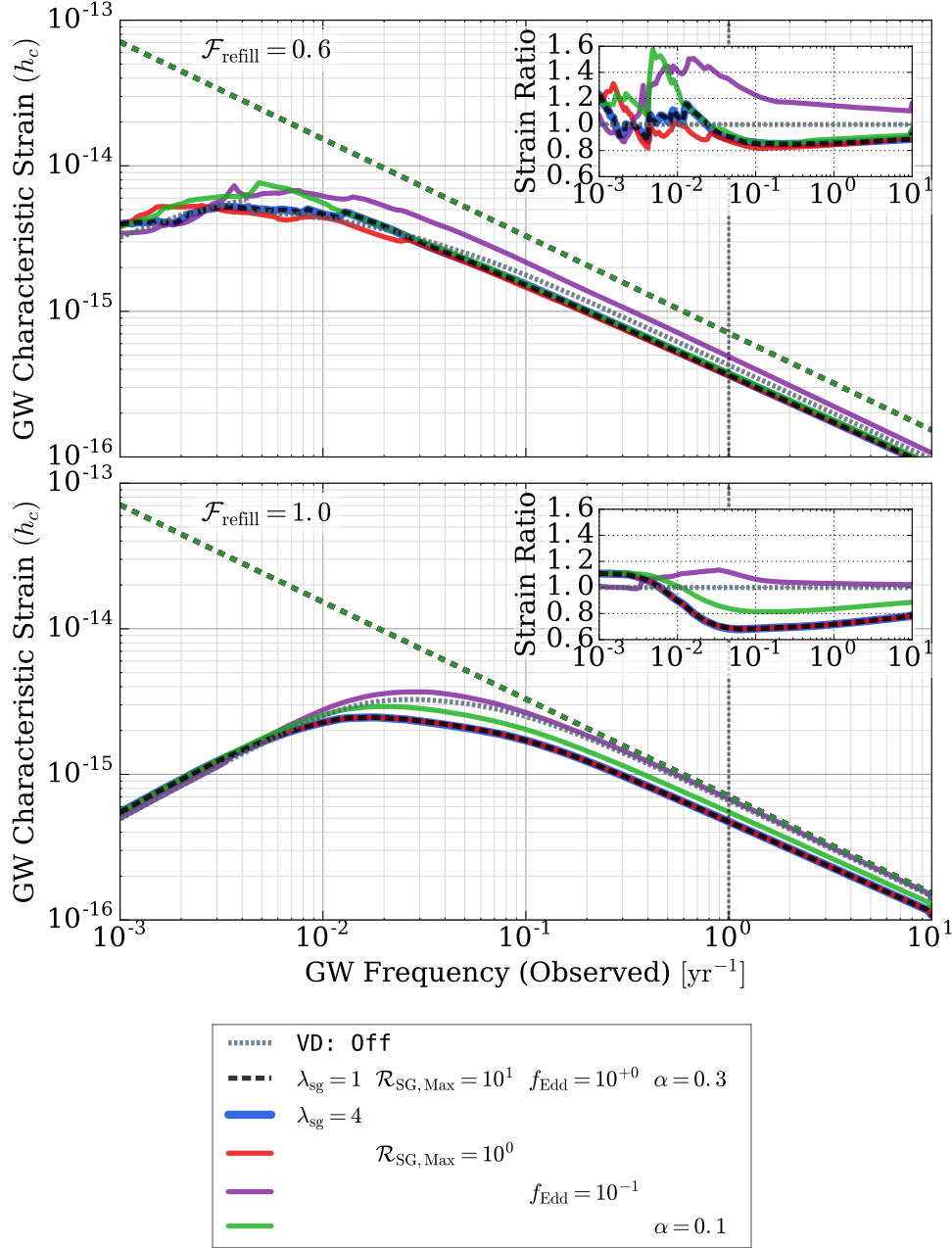
#### 1.4.2.4 EFFECTS OF CIRCUMBINARY VISCOUS DRAG ON THE GWB

The effects of viscous drag (VD) from a circumbinary disk are more subtle than those of DF and LC. Fig. 1.19 compares the GWB from our fiducial model (black, dashed) with a variety of VD parameter modifications (colored lines) and to a simulation with VD turned off (grey, dotted). All of these models use the ‘Enh-Stellar’ DF, but because there is virtually no overlap in between the VD and DF regimes, the results are very similar. The same VD parameters are explored as in the hardening rates shown in Fig. 1.9: modifying the self-gravity (SG) instability radius ( $\lambda_{\text{sg}}$ ), the maximum SG radius ( $\mathcal{R}_{\text{SG,Max}}$ ), the maximum accretion rate ( $f_{\text{Edd}}$ ), and the alpha viscosity parameter ( $\alpha$ ). The upper panel shows a moderately refilled LC ( $\mathcal{F}_{\text{refill}} = 0.6$ ), while in the lower panel the LC is always full ( $\mathcal{F}_{\text{refill}} = 1.0$ ). The inset panels show the ratio of GWB strain from each model to that of a ‘VD: Off’ (i.e. no disk) model.

The overall shape of the GWB spectrum and the location of the spectral turnover is again determined almost entirely by the LC. The circumbinary disk does affect an additional 10 – 40% amplitude modulation, tending to increase the amplitude at low frequencies ( $\lesssim 10^{-2} \text{ yr}^{-1}$ ) and decrease it at higher frequencies ( $\gtrsim 10^{-1} \text{ yr}^{-1}$ ). This reflects the same tradeoff between bringing more MBHB into each frequency band, versus driving them more rapidly through them. In the moderately (completely) refilled LC case, our fiducial VD model amounts to a  $\sim 20\%$  ( $\sim 30\%$ ) decrease in  $A_{\text{yr}^{-1}} = h_c(f = 1 \text{ yr}^{-1})$  and similarly at  $h_c(f = 10^{-1} \text{ yr}^{-1})$ .

At frequencies near the PTA band, the relationship between  $\mathcal{F}_{\text{coal}}$  and the GWB amplitude can be non-monotonic for VD variations, like with variations to the LC at low frequencies. For example, a comparison of Figures 1.9 & 1.19 shows that the  $\alpha = 0.1$  (green) model has the lowest fraction of high mass-ratio coalescences (with  $\mathcal{F}_{\text{coal}} = 0.22$ , versus  $\mathcal{F}_{\text{coal}} = 0.24$  for the fiducial model, and  $\mathcal{F}_{\text{coal}} = 0.31$  for the  $f_{\text{Edd}} = 0.1$  case) but an intermediate  $A_{\text{yr}^{-1}}$ .

One striking feature of the GWB strain ratios is the clear variations in spectral index, even at high frequencies. This is especially true for the always full LC, where the slope of the GWB can deviate by almost 10% from the canonical  $-2/3$  power-law. The disk-less model (grey, dotted) deviates by about 4% (3%) for  $\mathcal{F}_{\text{refill}} = 0.6$  ( $\mathcal{F}_{\text{refill}} = 1.0$ ) at  $f = 1 \text{ yr}^{-1}$ , due to a combination of residual LC scattering effects and some binaries stopping emitting after coalesce at varying critical frequencies. In our fiducial model (black, dashed), the deviations are more significant at 6% (8%). As different parameters make VD hardening more important at this frequency, the GWB amplitude decreases, and the spectral index tends to flatten. Our fiducial VD model tends to have among the strongest spectral deviations. Towards lower frequencies, where PTA are heading, the turnover in the GWB spectrum becomes more significant, especially if the LC is effectively refilled. At  $f = 10^{-1} \text{ yr}^{-1}$ , for example, our fiducial model (‘Enh-Stellar’,  $\mathcal{F}_{\text{refill}} = 0.6$ ) gives a spectral index of about  $-0.6$ , while for  $\mathcal{F}_{\text{refill}} = 1.0$  it becomes slightly flatter than  $-0.4$ . A sum-



**Figure 1.19:** Gravitational wave background from varying viscous drag (VD) parameters. Simulations with a variety of VD models are compared, with our fiducial model in dashed black, a model with no-VD in dotted grey, and each color of line showing changes to a different parameter. For comparison, the power-law model using all Illustris MBHB is also shown. The parameters modified are: the self-gravity (SG) instability radius ( $\lambda_{\text{sg}}$ ), the maximum SG radius ( $\mathcal{R}_{\text{SG,Max}}$ ), the maximum accretion rate ( $f_{\text{Edd}}$ ), and the alpha viscosity parameter ( $\alpha$ ). The hardening rates for each of these models are shown in Fig. 1.10. The upper and lower panels show simulations for different LC refilling fractions. The inset panels show the ratio of GWB amplitude from each model to the ‘VD: Off’ case, as a function of GW frequency. Different VD parameters change  $A_{\text{yr}^{-1}}$  by 10 – 40%, and the spectral slope at  $1 \text{ yr}^{-1}$  by up to  $\sim 10\%$ .

mary of GWB amplitudes and spectral indices are presented in Table 1.3, for a variety of configurations.

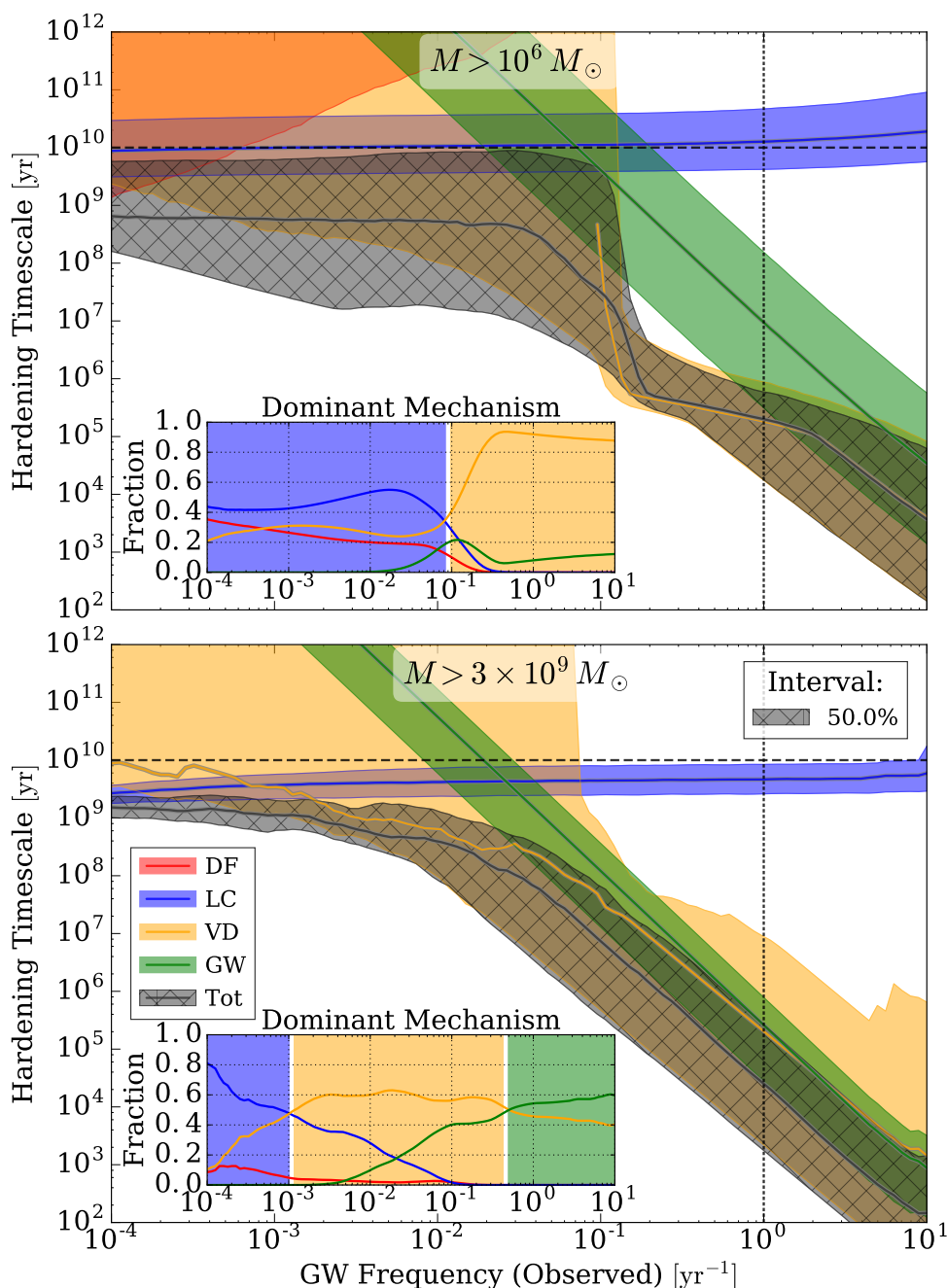
For a given binary system, GW radiation will *always* dominate at some sufficiently small separation (high frequency) where the circumbinary disk dynamically decouples from the hardening MBHB. This does *not necessarily* mean, however, that after considering a full ensemble of MBHB systems, with a variety of masses, that there is any frequency band with a spectral slope identical to the purely GW-driven case ( $h_c \propto f^{-2/3}$ ). Hardening rates as a function of GW frequency are shown in Fig. 1.20. The upper panel includes all MBHB in our sample\*, for which we see that VD remains dominant well above the PTA frequency band. The high total-mass systems ( $M > 3 \times 10^9 M_\odot$ )—which contribute the bulk of the GWB signal—are shown in the lower panel. These binaries tend to be driven in roughly equal amounts by VD and GW hardening at the frequencies where PTA detections should be forthcoming.

Figure 1.20 (and Fig. 1.11) show that the typical hardening rates for VD are very similar to that of GW radiation. Indeed, as discussed in §1.3.3, the inner-most disk region has hardening times  $\tau_{v,1} \propto r^{7/2}$ , while that of purely gravitational wave emission is  $\tau_{gw} \propto r^4$ . Hardening rates for farther-out disk regions tend to deviate more strongly from that of purely GW evolution, which could become more important for lower density disks.

The MBH accretion rates, which set the density of the circumbinary disks in our models, are perhaps one of the more uncertain aspects of the Illustris simulations, given that the accretion disk scale is well below the resolution limit and must therefore rely on a sub-grid prescription. Additionally, out of all possible configurations, the fiducial disk parameters we adopt tend to produce fairly strong effects on the GWB. If, for example, a  $\beta$ -disk model is more accurate, or the  $\alpha$ -viscosity should be lower, the effects in the PTA band will be more moderate (see, e.g., Kocsis & Sesana 2011). None the less, we consistently see GWB spectral indices between  $-0.6$  and  $-0.65$  at  $1 \text{ yr}^{-1}$ , for a wide variety of model parameters. While these  $\lesssim 10\%$  deviations may be entirely unobservable in PTA observations (especially after taking stochastic variations into account; e.g. Sesana, Vecchio & Colacino 2008), it may need to be considered when using priors or match-filtering for detecting a GWB. More stringent observational constraints on specifically post galaxy-merger AGN activity could be used to better calibrate the VD model.

---

\*Recall that we select only MBH with masses  $M > 10^6 M_\odot$ .



**Figure 1.20:** Binary hardening timescales versus GW frequency, by mechanism, for our fiducial model ( $\mathcal{F}_{\text{refill}} = 0.6$ , DF: ‘Enh-Stellar’). Lines and bands show the median and 50% intervals for individual mechanisms (colors), along with the total hardening rate (grey, hatched). The inset panels show the fraction of binaries dominated by each mechanism, again versus frequency. The upper panel shows all MBHB in our sample, while the lower panel includes only systems with total mass above  $3 \times 10^9 M_{\odot}$ , roughly where the bulk of the GWB amplitude comes from. Only for the high mass systems do the majority become dominated by GW emission at high frequencies, with VD still contributing substantially to the overall hardening timescales.

### 1.4.3 THE POPULATIONS OF MBHB

For the first time, we have used cosmological, hydrodynamic models which self-consistently evolve dark matter, gas, stars and MBH, to more precisely probe the connection between MBHB mergers and their environments. Previous calculations (see §1.1) of the GWB using SAM prescribe MBH onto their galaxies based on scaling relations. The MBH population in Illustris, on the other hand, co-evolves with, and shapes, its environment. These data are then much better suited to analyze the details of MBHB and GW source populations, and their hosts.

Figure 1.21 shows the distribution of properties for sources contributing to the GWB, from top to bottom: total mass ( $M$ ), mass ratio ( $\mu$ ), and redshift ( $z$ ). In the left column, these properties are weighted by squared-strain\* for each source, and the resulting one-, two-, and three-sigma contours are shown as a function of GW frequency. The right column shows the cumulative distribution over the same source properties, weighted by  $A_{\text{yr}^{-1}}^2$  (solid), compared to the unweighted distribution of all sources contributing at  $f = 1 \text{ yr}^{-1}$ . Strain-weighted sources tend to be at higher mass-ratios and much higher masses. While the fraction of all binaries rises fairly smoothly with total masses between  $10^7$  and  $10^9 M_\odot$  (dashed, black line; top-right panel), 90% of the GWB is contributed (solid, black line) by binaries with total mass  $\gtrsim 10^9 M_\odot$ —simply showing the strong dependence of the GW strain on the total system mass.

The core contribution over all three parameters tends to remain fairly constant over GW frequency, with median values around  $M \approx 4 \times 10^9 M_\odot$ ,  $\mu \approx 0.3$ , and  $z \approx 0.3$ . The tails of the distribution drop to noticeably lower values when moving to higher frequencies. This is especially pronounced in the redshift distribution, where at frequencies of a few times  $10^{-3} \text{ yr}^{-1}$  virtually all GWB-weighted sources come from  $z > 10^{-2}$ , while at  $f = 1 \text{ yr}^{-1}$ , almost 10% are below that redshift. While  $\sim 20\%$  of binaries that reach  $f = 1 \text{ yr}^{-1}$  come from redshift above  $z = 1$ , they only contribute  $\sim 0.5\%$  of the GWB amplitude. Lower redshift and higher mass-ratio systems do contribute somewhat disproportionately to the GWB amplitude, but their distributions are altogether fairly consistent with the overall population. The presence of a non-negligible fraction of low redshift sources motivates the need to explore populations of MBHB in the local universe which could be resolvable as individual ‘stochastic’ sources, or contribute to angular anisotropies in the GW sky. An analysis of our results in this context is currently underway, and the results will be presented in a future study.

As we move into the forthcoming era of PTA *detections* it will be increasingly important to use self-consistent hydrodynamic models to better understand the coupling of the MBH populations to their host-galaxies and merger environments. The Illustris host-galaxy properties of our MBHB, at the time of binary formation, are presented in Fig. 1.22. We show stellar radius, stellar mass and ‘subhalo mass’, and each of these properties† is strongly biased towards higher values when weighting by GW strain. In particular, the median, strain-weighted subhalo and stellar masses are each more than an order of magnitude larger than the median of the host-galaxy population by number. The bias is exceedingly strong for stellar mass, where  $\sim 90\%$  of the GWB amplitude is contributed by only  $\sim 20\%$  of MBHB host galaxies.

Following the galaxies that host MBHB to observe their parameters at the times they contribute significantly to the GW spectrum will be important for any future multi-messenger observations using PTA or predicting and deciphering anisotropies in the GWB (Taylor & Gair 2013; Mingarelli et al. 2013; Taylor et al. 2015). Better understanding host galaxy properties as they evolve in time could also be useful in

\*As seen in Eq. 1.18, binaries contribute to the strain spectrum in quadrature.

†The stellar radius is measured as the stellar half-mass radius ( $R_{\star,1/2}$ ); the stellar mass is the mass of star particles within  $R = 2R_{\star,1/2}$ ; and the subhalo mass is the combined mass of all particles and cells associated with the host galaxy. While these simulation measurements are, of course, non-trivial to relate to their observational counterparts, they are useful for relative comparisons.

understanding whether ‘offset’ AGN (those distinctly separated from the morphological or mass-weighted center of their galaxies) are due to binarity (i.e. a recent, or perhaps not so recent, merger) or possible due to post-coalesce GW ‘kicks’ (e.g. Blecha et al. 2016).

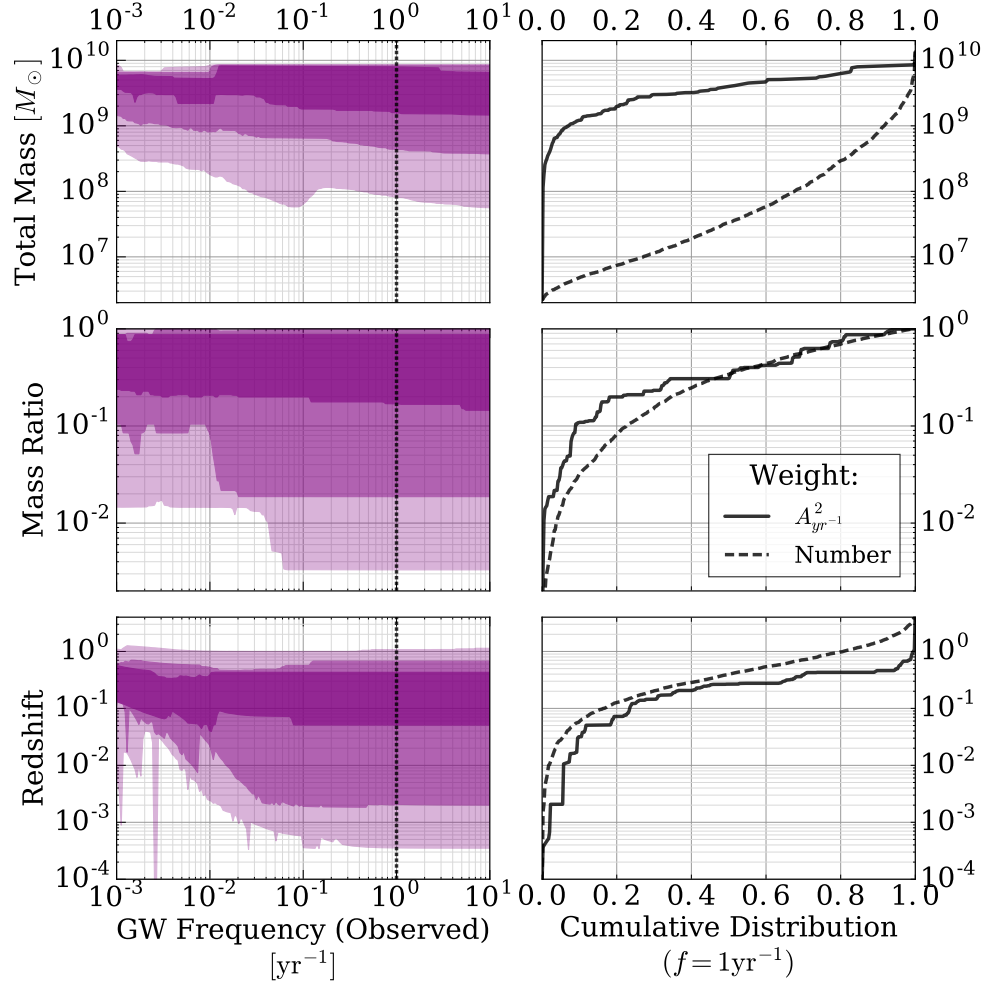
Of great observational interest is the presence (e.g. Comerford et al. 2015), or perhaps conspicuous absence (e.g. Burke-Spolaor 2011), of dual and binary AGN. The observational biases towards finding or systematically excluding MBH binaries with electromagnetic observations are extremely complex. None the less, understanding the characteristic residence times of binaries at different physical separations, the types of host galaxies they occupy, and the probability they will be observable (e.g. via the amount of gas available to power AGN activity) is crucial to backing out the underlying population, and placing empirical constraints on models of MBHB inspiral. A systematic study of this topic using these data is currently underway (Kelley et al. in prep.).

The fraction of MBHB which persist (i.e. remain uncoalesced) at redshift zero are shown in Fig. 1.23 as a function of total mass (left column) and mass ratio (right column). Three different separation criteria are shown in each panel:  $r > 0.0$  (i.e. any persisting MBHB; dark, dotted),  $r > 1$  pc (medium, dot-dashed), and  $r > 10^2$  pc (light, dashed). Each row corresponds to a different DF model, and line colors vary by LC refilling fractions. In general, persisting fractions fall rapidly with increasing total mass and moderately with increasing mass ratio, until nearly equal-mass systems where the persisting fractions plummet.

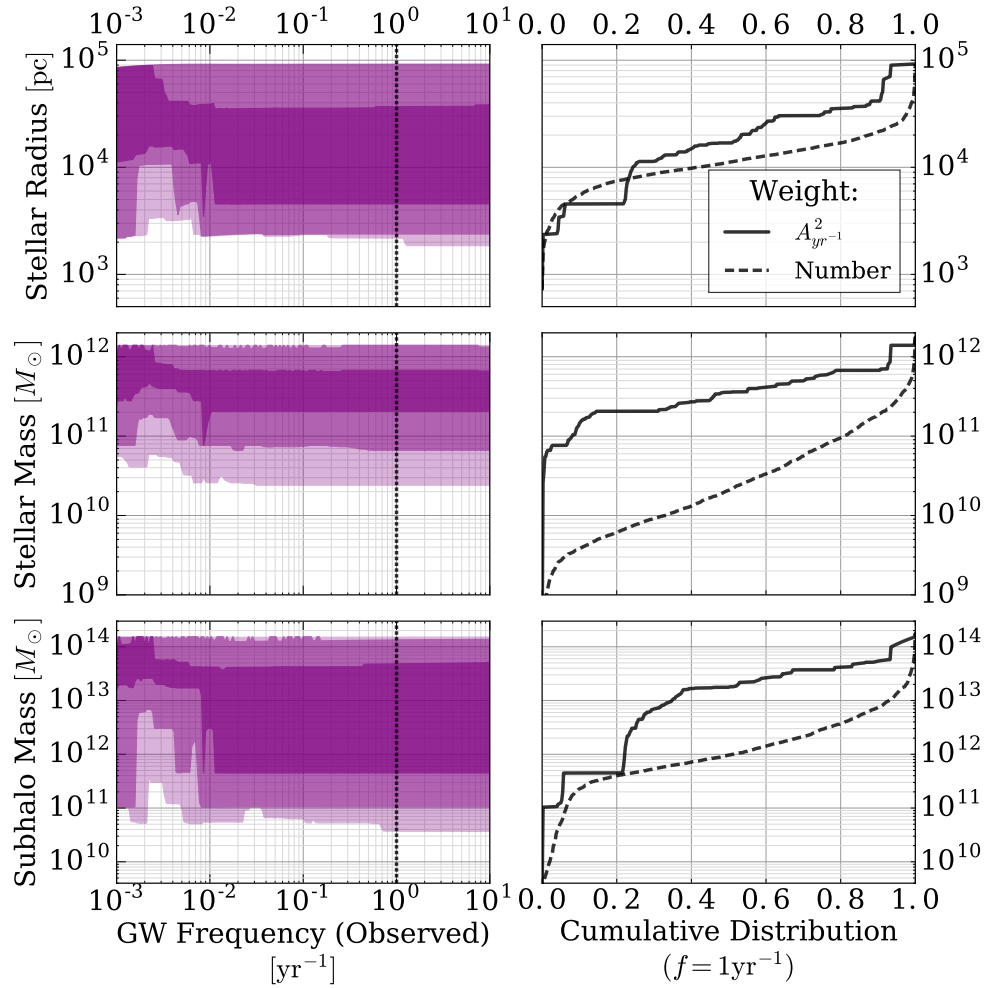
The specific persisting fraction depends quite sensitively on both  $\mathcal{F}_{\text{refill}}$  and DF model. The ‘Enh-Stellar’ model has by far the most persisting systems, and relatively slight variance with either separation criteria or  $\mathcal{F}_{\text{refill}}$ . For our fiducial model with  $\mathcal{F}_{\text{refill}} = 0.6$ , 80% of all binaries persist, with only weak trends with either total mass or mass ratio: 77% with  $M > 10^8 M_{\odot}$ , and 74% with  $\mu > 0.2$ . For systems fulfilling both requirements, the persisting fraction drops more noticeably to 55%.

The  $r > 10^2$  pc population in particular, is almost solely determined by DF, as the other hardening mechanisms take effect only at smaller scales. At these large separations, persisting fractions for our fiducial model are 46% (all), 45% ( $M > 10^8 M_{\odot}$ ), and 33% ( $\mu > 0.2$ ), but for both high total-mass and mass-ratio, the widely-separated persisting fraction drops dramatically to only 1%. If the DF is more effective, as in the ‘Enh-Gyr’ model, these fractions decrease significantly to 11% (all), 15% ( $M > 10^8 M_{\odot}$ ), 6% ( $\mu > 0.2$ ), and 1% ( $M > 10^8 M_{\odot} \& \mu > 0.2$ ). A summary of persisting fractions at both  $r > 10^2$  pc and  $r > 1$  pc, for mass combinations and DF & LC models are presented in Table 1.3.

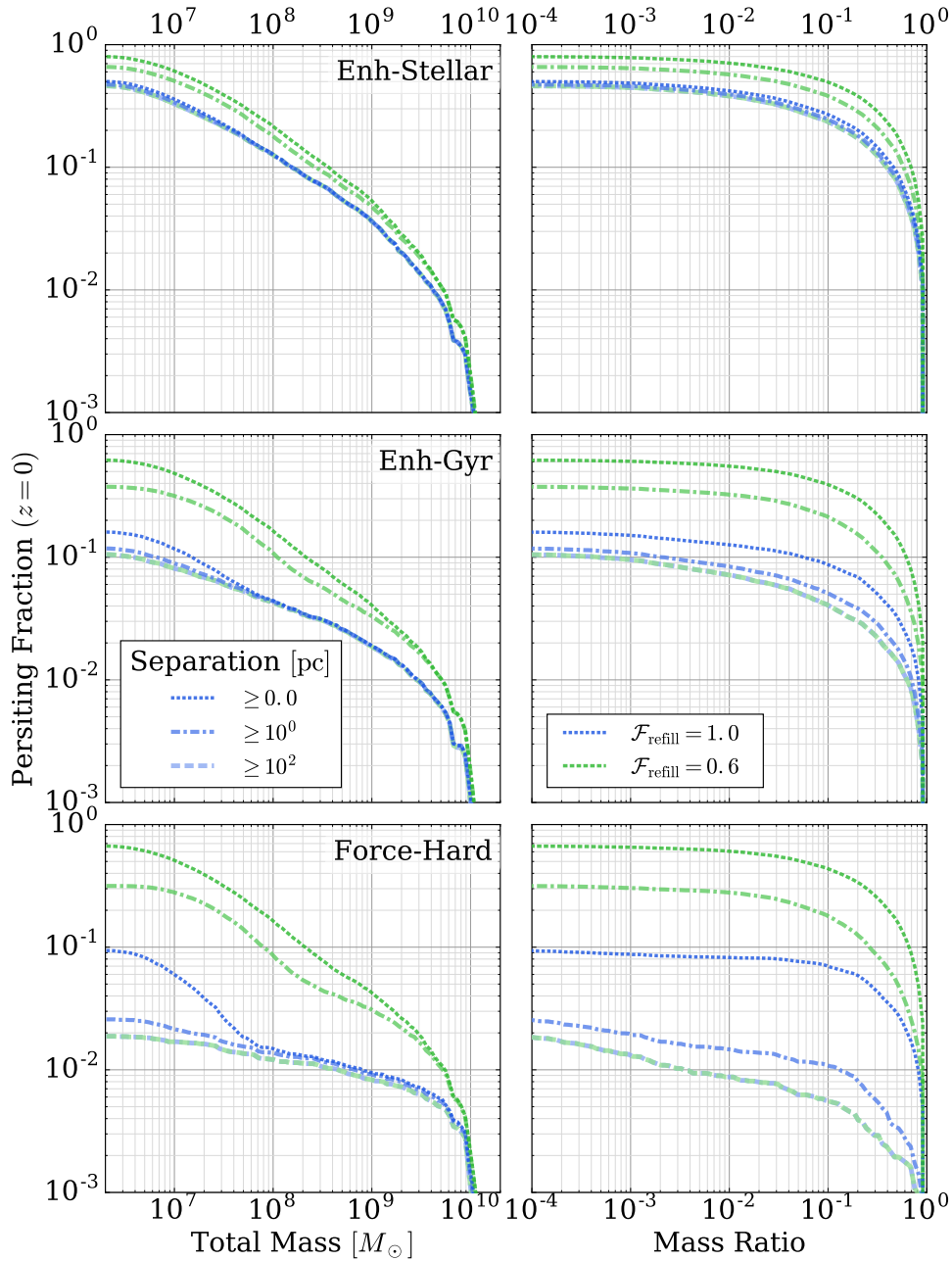




**Figure 1.21:** Population of MBHB contributing to the GWB. The left column shows, from top to bottom, the MBHB total mass, mass ratio, and redshift, weighted by each system’s contribution to the GWB amplitude. Contours represent one-, two-, and three- sigma intervals. The Right column shows cumulative distributions, at a frequency of  $1 \text{ yr}^{-1}$ , for the same parameters. The solid line weights by contribution to the GWB amplitude ( $A_{\text{yr}^{-1}}^2$ ) and the dashed line is the distribution of the number of sources *contributing at  $1 \text{ yr}^{-1}$* . The median values by GWB-contribution are roughly constant over GW frequency, at  $M \approx 4 \times 10^9 M_\odot$ ,  $\mu \approx 0.3$ , and  $z \approx 0.3$  for this, our fiducial model. The overall distribution of sources moves noticeably to include lower masses, mass ratios, and redshifts at higher GW frequencies. The contribution from redshifts above  $z \approx 0.4$  drops sharply, with  $\lesssim 1\%$  of the GWB signal coming from  $z > 1.0$ , while still  $\sim 20\%$  of all binaries emit there.



**Figure 1.22:** The properties of host galaxies for the population of MBHB contributing to the GWB. From top to bottom, rows show the stellar (half-mass) radius, stellar mass (within twice the stellar half-mass radius), and subhalo mass (mass of all particles associated with the galaxy). The left column shows these parameters weighted by their resident MBHB's contribution to the GWB as a function of frequency. The right column shows the cumulative distribution at  $f = 1 \text{ yr}^{-1}$ , both for contribution to GWB amplitude (solid) and by overall number (dashed). The GWB comes from MBHB predominantly in galaxies which are over-sized and significantly over-massive—especially in stars.



**Figure 1.23:** Fraction of binaries which persist at redshift zero as a function of total mass (left column) and mass ratio (right column). Three different DF models are compared, from top to bottom: ‘Enh-Stellar’, ‘Enh-Gyr’, and ‘Force-Hard’; and in each case, LC refilling parameters of  $\mathcal{F}_{\text{refill}} = 0.6$  (green) and  $\mathcal{F}_{\text{refill}} = 1.0$  (blue) are compared. Different line patterns show binaries with different separations: all separations ( $r \geq 0.0$ , dotted),  $r \geq 1$  pc (dot-dashed), and  $r \geq 10^2$  pc (dashed). Note that in each panel the  $r \geq 10^2$  pc distributions are indistinguishable between LC parameters as the LC only takes effect at and below about  $10^2$  pc. The fraction of persisting systems is very strongly dependent on both DF and LC model. For ‘Enh-Stellar’, the most conservative DF case, a large fraction of systems remain in the DF regime ( $r \sim 10^2$  pc), before LC scattering can have a significant effect. ‘Force-Hard’, on the other hand, represents an approximately optimal DF at large scales, and shows a corresponding dearth of wide separation systems. Observations of the true fraction of systems at these separations could strongly constrain the efficiency of these hardening mechanisms.

#### 1.4.4 MBH TRIPLES

The long characteristic lifetimes we see in our MBHB populations, and the (at times) substantial number of systems which remain at large separations, immediately begs the question of how often a third MBH (i.e. second galaxy merger) could become dynamically relevant. For  $\mathcal{F}_{\text{refill}} = 1.0$ , the median lifetimes of our MBHB tend to be comparable to the median time between binary formation events, and for  $\mathcal{F}_{\text{refill}} = 0.6$ , they are almost an order of magnitude longer. After selection cuts (see §1.2), 37% of our MBH binaries have subsequent ‘merger’ events (i.e. a second ‘merger’ is recorded by Illustris involving an MBH meeting our selection criteria). In our implementation, each of those binary systems are evolved completely independently, even if parts of their evolution are occurring simultaneously\*. With this caveat, we can still consider, very simplistically, in how many systems the *second* binary overtakes the first as they harden. Out of the binaries with subsequent events, 83% (31% of all binaries) are overtaken, 76% (28% overall) of those before redshift zero, and 42% (16%) with  $z > 0.0$  and mass ratio  $\mu > 0.2$ .

The tendency for subsequent binaries to cross in our simulations likely reflects systems’ ability to increase noticeably in total mass over the course of the merger process. We emphasize that this is a very simplistic and preliminary investigation. If, for example, MBH remnants tend to receive significant ‘kicks’ after merger, the resulting fractions could change significantly. None the less, the apparent commonality of candidate multiples suggests that the role of triples should be investigated more thoroughly.

It is unclear how such triple systems should be treated, even in a simple semi-analytic manner (see, however, [Bonetti et al. 2016](#)). The conventional wisdom of triple system dynamics is that the lowest mass object will be ejected, while the more massive pair become bound in a binary (e.g. [Hills 1975](#)). Such ‘exchange’ interactions are motivated primarily from stochastic scattering events, like those which may occur between stars in dense stellar environments. In these cases, the system can be viewed as nearly dissipationless, and their initial encounter is effectively stochastic. It is our premise, however, that the environments and dynamics of MBH multiples are heavily dissipational. For example, consider an initial pair of MBH which encounter at kpc scales, on a hyperbolic orbit. If the system quickly circularizes, and hardens to scales of 1 – 100 pc, then a third MBH which encounters the system—again at kpc scales—may similarly settle into an outer, roughly-circular orbit forming a hierarchical system. In such a situation, secular instead of scattering dynamics, such as the Kozai-Lidov mechanism ([Lidov 1962](#); [Kozai 1962](#)) or resonant migration may be more appropriate than traditional three-body scatterings. In this case, the outer MBH in the triple system may accelerate the hardening of the inner binary, driving it to coalescence (e.g. [Blaes, Lee & Socrates 2002](#)). This may be less likely in gas-rich environments which could effectively damp eccentric evolution, but here gas-driven inspiral will likely cause rapid coalescence in any case.

MBH triples forming hierarchically with low to moderate eccentricities may evolve in a resonant fashion. If, on the other hand, environmental effects sufficiently enhance (or preserve initially high) eccentricities of MBHB, then the resulting highly radial orbits may strongly intersect. In that case, a more stochastic-scattering-like regime may indeed still be appropriate. Numerous studies have suggested that environmental effects can indeed enhance MBHB eccentricity (e.g. [Quinlan 1996](#); [Sesana 2010](#); [Ravi et al. 2014](#)). If the interaction between MBH triples (or even higher-order multiples) is indeed most similar to scattering, then the simplest prescription of removing the lowest-mass BH, with or without some additional hardening of the more massive pair, may still be appropriate (e.g. [Hoffman & Loeb 2007](#)). An ejected MBH which may later fall back to the galactic center, while of great observational interest in and of itself, is likely less important for GW emission *per se*.

The observation of a triple-AGN system could provide insight into the type of system they form

\*Recall that in Illustris, after the initial ‘merger’ event, only a single remnant MBH particle remains.

(i.e. hierarchical vs. scattering), and their lifetimes. Additionally, MBH ejected by three-body interactions could be observable as offset AGN, and possible confused with binary MBH, or ones ‘recoiling’ from previous coalescences (e.g. Blecha et al. 2011). We have assumed that recoiling systems do not significantly affect our populations, effectively assuming that kicks are small—which is expected for spin-aligned MBHB. This is motivated by studies which have shown that gravitational torques from circumbinary disks, such as those we consider, can be effective at aligning spins on timescales significantly shorter than a viscous time (e.g. Bogdanović, Reynolds & Miller 2007; Dotti et al. 2010; Miller & Krolik 2013).

The MBH populations from the Illustris simulations are well suited for this problem, as they accurately follow the histories and large-scale environments of MBHB systems and host galaxies. As we are currently working on implementing eccentric evolution into our simulations, we plan to explore multi-MBH systems in more detail. This framework will also allow for the treatment of kicked MBH resulting from random spin orientations, if for example the spins of a substantial fraction of MBHB occur in gas-poor environments in which they may not be aligned.

## 1.5 CONCLUSIONS AND SUMMARY

For the first time, we have used the results of self-consistent, hydrodynamic cosmological simulations, with a co-evolved population of Massive Black Holes (MBH) to calculate the plausible stochastic Gravitational-Wave Background (GWB) soon to be detectable by Pulsar Timing Arrays (PTA). We have also presented the first simultaneous, numerical treatment of all classes of MBH Binary (MBHB) hardening mechanisms, discussing the effects of each: dynamical friction, stellar (loss-cone) scattering, gas drag from a viscous circumbinary disk, and gravitational wave emission.

The most advanced previous studies have included only individual environmental effects, for example, calculating dynamical friction (DF) timescales to determine which systems will contribute to the GWB (McWilliams, Ostriker & Pretorius 2014), or attenuating the GWB spectrum due to loss-cone (LC) stellar-scattering (Ravi et al. 2014). We explicitly integrate each of almost ten thousand MBH binaries, from galactic scales to coalescence, using self-consistently derived, realistic galaxy environments and MBH accretion rates. We thoroughly explore a broad parameter space for each hardening mechanism to determine the effects on the MBHB merger process, the lifetimes of systems, and the resulting GWB spectrum they produce.

The resulting lifetimes of MBHB that coalesce by redshift zero are usually gigayears, while that of low total-mass and extreme mass-ratio systems typically extend well above a Hubble time. In our fiducial model, with a modest DF prescription (‘Enh-Stellar’) and moderately refilled LC ( $\mathcal{F}_{\text{refill}} = 0.6$ ), the median lifetime of MBHB with total masses  $M > 10^8 M_{\odot}$  is 17 Gyr, with 23% coalescing before redshift zero. Massive systems that also have high mass ratios,  $\mu > 0.2$ , merge much more effectively, with a median lifetime of 6.9 Gyr and 45% coalescing at  $z > 0$ . Increasing the effectiveness of the LC drastically decreases system merger times. For an always full LC ( $\mathcal{F}_{\text{refill}} = 1.0$ ), the lifetime of massive systems decreases to 4.9 Gyr and 0.35 Gyr for systems with  $M > 10^8 M_{\odot}$  and all mass ratios, and those with  $\mu > 0.2$  respectively. The coalescing fractions in these cases doubles to 54% and 99%. A summary of lifetimes and coalescing fractions for different models is presented in Table 1.3.

The growing number of dual-MBH candidates (e.g. Deane et al. 2014; Comerford et al. 2015) presents the opportunity to constrain binary lifetimes and coalescing fractions observationally. For most of our models, only about 1% of MBHB with total masses  $M > 10^8 M_{\odot}$  and mass  $\mu > 0.2$  remain at separations  $r > 10^2$  pc at redshift zero. At smaller separations,  $r > 1$  pc, the fractions are dependent on model parameters, but in general between 1 – 40%. Tabulated persisting fractions are included in Table 1.3 for

a variety of models and situations. Observational constraints on these fractions can narrow down the relevant parameter space of hardening physics. Accurate predictions for dual-MBH observations must fold in AGN activity fractions and duty cycles, and their correlations with binary merger lifetimes. A comprehensive study of dual-AGN observability predicted by our models, over redshift and different observational parameters, is currently underway.

In addition to measuring the fraction of MBH in associations (e.g. dual AGN) as a function of separation, the redshift distribution of dual-MBH can be useful in understanding their evolution. The Illustris simulations, for example, give a median MBHB formation redshift\* of  $z \approx 1.25$ . Depending on the parameters of the hardening models, the median coalescing redshift can be anywhere between  $z \approx 0.4 - 1.0$ , with  $z \approx 0.6$  suggested by our fiducial model.

Without electromagnetic observations, GWB detections and upper limits can also be used to inform our understanding of MBH evolution (e.g. [Sampson, Cornish & McWilliams 2015](#)). Even if the fraction of systems which coalesce is quite low, the most massive and high mass-ratio systems, which produce the strongest GW, are difficult to *keep* from merging. In a simulation with the weakest hardening rates (‘Enh-Stellar’;  $\mathcal{F}_{\text{refill}} = 0.0$ ) only  $\sim 12\%$  of all binaries coalesce by redshift zero, but the GWB amplitude at  $f = 1 \text{ yr}^{-1}$  is still  $0.2 \times 10^{-15}$ —only about a factor of five below the most recent upper limits<sup>†</sup>.

In our fiducial model, we use a moderate LC refilling rate ( $\mathcal{F}_{\text{refill}} = 0.6$ ) which increases the number of MBHB contributing to the GWB at  $1 \text{ yr}^{-1}$ , producing an amplitude of  $A_{\text{yr}^{-1}} \approx 0.4 \times 10^{-15}$ . Increasing the effectiveness of DF and/or LC scattering tends to increase the amplitude further. Our fiducial model also includes fairly strong viscous drag (VD) from circumbinary disks, which decreases the time MBHB spend emitting in each frequency band, and thus attenuating the GWB. This effect tends to be more subtle, producing GWB attenuation of about 15%. In general, for a fairly broad range of parameters, our simulations yield GWB amplitudes between  $\sim 0.3 - 0.6 \times 10^{-15}$ . A GWB amplitude of  $A_{\text{yr}^{-1}} \approx 0.4 \times 10^{-15}$  is less than a factor of three below current detection limits—a parameter space which will likely be probed by PTA within the next decade.

The most stringent PTA upper limits of  $A_{\text{yr}^{-1}} \lesssim 10^{-15}$  ([Shannon et al. 2015](#)) have already excluded a broad swath of previous predictions. Many of those models assume that binary hardening is very effective, with all MBHB quickly reaching the PTA band and emitting an unattenuated signal—i.e. evolving purely due to GW-emission, without additional environmental hardening effects. Following the same procedure, to calculate an upper-limit to the GWB based on our population of MBHB, we find a GWB amplitude of  $A_{\text{yr}^{-1}} \approx 0.7 \times 10^{-15}$ —slightly below the PTA limit. The Illustris simulation volume is very large for a hydrodynamic simulation, but it lacks the very-rare, most massive MBH in the universe ( $\gtrsim 10^{10} M_{\odot}$ ) which could slightly increase our predicted GWB amplitude—although, likely a correction on the order of  $\sim 10\%$ <sup>‡</sup> (Sesana, private communication). None the less, our upper limit suggests that the current lack of PTA detections shouldn’t be interpreted as a missing signal.

Our upper-limit value of  $A_{\text{yr}^{-1}} \approx 0.7 \times 10^{-15}$  falls just within the lower end of some recent studies (e.g. [Ravi et al. 2014](#); [Roebber et al. 2016](#)), but is generally lower than much of the previous literature (see e.g. Table 1.1, and Fig. 2 of [Shannon et al. 2015](#)). Likely, this is at-least partly because the MBH merger

\*Recall that MBHB ‘formation’, in this context, corresponds to two MBH coming within a few kpc of each other.

<sup>†</sup>Previous studies have shown that including eccentric evolution can significantly decrease the GWB amplitude ([Ravi et al. 2014](#)), so we caution that the weakest GWB observed in our simulations, which *do not* include eccentric evolution, may not be a robust lower limit. We are currently exploring the effects of eccentricity, and altered MBH-Host scaling relations on the minimum plausible GWB—to be presented in a future study.

<sup>‡</sup>The effects of simulation volume on the predicted GWB amplitude should be studied more carefully to confirm this estimate.

rates derived from Illustris are based directly on simulated galaxy-galaxy merger rates. The bulk of existing calculations have either used inferences from (dark matter only) halo-halo mergers which may have systematic issues (see, e.g., [Rodriguez-Gomez et al. 2015](#)), or observations of galaxy merger rates which have uncertain timescales. This upper-limit is based on optimistic, GW-only evolution. In our fiducial model, the signal is lower by  $\sim 50\%$  due primarily to the moderately refilled LC, and mildly due to VD attenuation.

Variations in the rate at which the stellar LC is refilled has the strongest effect on the shape and amplitude of the GWB spectrum in our simulations, especially at low frequencies. PTA observations are moving towards these frequencies, as the duration of their timing baselines increase. Unlike at higher frequencies where scattering increases the number of MBHB contributing to the GWB, at  $10^{-1} \text{ yr}^{-1}$ , for example, effective LC refilling leads to attenuation of the GWB from accelerated binary hardening. Here, our spectra tend to lie at amplitudes between  $1.5 - 2.5 \times 10^{-15}$ , with spectral indices between about  $-0.4$  and  $-0.6$ —a significant deviation from the canonical  $-2/3$  power-law. At frequencies lower still ( $f \lesssim 10^{-2} \text{ yr}^{-1}$ ), the effective LC scattering produces a strong turnover in the GWB spectrum. A summary of GWB amplitudes and spectral indices is presented in [Table 1.3](#) for both  $f = 1 \text{ yr}^{-1}$  and  $10^{-1} \text{ yr}^{-1}$ , and numerous hardening models.

In our fiducial simulation, we find that the median contribution to the GWB comes from binaries at a redshift of  $z \approx 0.3$ , with total masses  $M \approx 10^9 M_{\odot}$ , and mass ratios  $\mu \approx 0.3$ . The co-evolved population of MBH and galaxies in Illustris allows us to also examine typical host-galaxy properties for the first time. Galaxies containing MBHB contributing strongly to the GWB are noticeably larger and more massive galaxies. The median stellar mass of galaxies, weighted by GWB contribution, is about  $3 \times 10^{11} M_{\odot}$ —more than an order of magnitude larger than the median stellar mass for all galaxies.

Based on the merger trees and binary lifetimes produced from our simulations, we have also shown that the presence of higher-order MBH multiples could be a non-negligible aspect of MBH evolution. The simplest examination suggests that triples could be important in about 30% of MBHB in our simulations. In future work, we will explore these triple systems in more detail, as well as the effects of nonzero eccentricity and post-merger recoils. We also hope to implement more self-consistent LC refilling, and more comprehensive tracking of the changing galactic environment.

In summary,

- **MBH binary lifetimes tend to be multiple Gyr, even for massive systems.** While massive and high mass ratio systems are likely rare at very large separations, observations of dual MBH at  $r > 1 \text{ pc}$  can be used to constrain the merger physics.
- **The GWB amplitude predicted by our models is  $A_{1 \text{ yr}^{-1}} \approx 0.4 \times 10^{-15}$ , with a range of about  $0.3 - 0.6 \times 10^{-15}$  for different hardening parameters.** At lower frequencies, we find  $A_{0.1 \text{ yr}^{-1}} \approx 1.5 - 2.5 \times 10^{-15}$ , with spectral indices between  $-0.4$  &  $-0.6$ —a noticeable deviation from the canonical  $-2/3$  power-law.
- **We find that the lack of PTA detections so far is entirely consistent with our MBH population,** and does not require environmental effects. At the same time, our most conservative models yield a GWB amplitude of  $A_{\text{yr}^{-1}} = 0.2 \times 10^{-15}$ . While incorporating non-zero eccentricities may further suppress our GWB predictions, our simulations suggest that if PTA limits improve by a factor of 3–4 and no detection is made, our understanding of galaxy and MBH evolution may require revision.
- **The median redshift and total mass of MBHB sources contributing to the GWB are  $z \approx 0.3$  and a few  $10^9 M_{\odot}$ ,** while the median coalescence time of all systems tends towards  $z \approx$

0.6. Observations of the redshift distribution and host galaxy properties of dual-MBH can be informative for our understanding of binary evolution.

- **Our simulations suggest that up to 30% of binaries could involve the presence of a third MBH.** The role of MBH triples is currently unclear, but should be explored and included in future simulations.

The environments around MBHB form a complex and interwoven parameter space with additive and often degenerate effects on the GWB. Better constraints on MBH–host correlations, combined with increasingly strict upper limits on the GWB amplitude, will soon tightly constrain the efficiency with which MBHB are able to coalesce. That efficiency then determines the fraction of galaxies which should have observable dual or binary AGN, providing an additional test of our most fundamental assumptions of MBH/galaxy growth and co-evolution. We believe that our results, and similar analyses, can be used to leverage GWB observations along with dual and offset AGN to comprehensively understand the MBH population and their evolution. These exotic binaries involve a plethora of dynamical processes which are still poorly understood, but affect our most fundamental assumptions of black holes in the universe and thus the evolution of galaxies over cosmological time. Right now, we are entering the era of GW astronomy, and with it, a direct view of BH physics and evolution on all scales.

#### ACKNOWLEDGMENTS

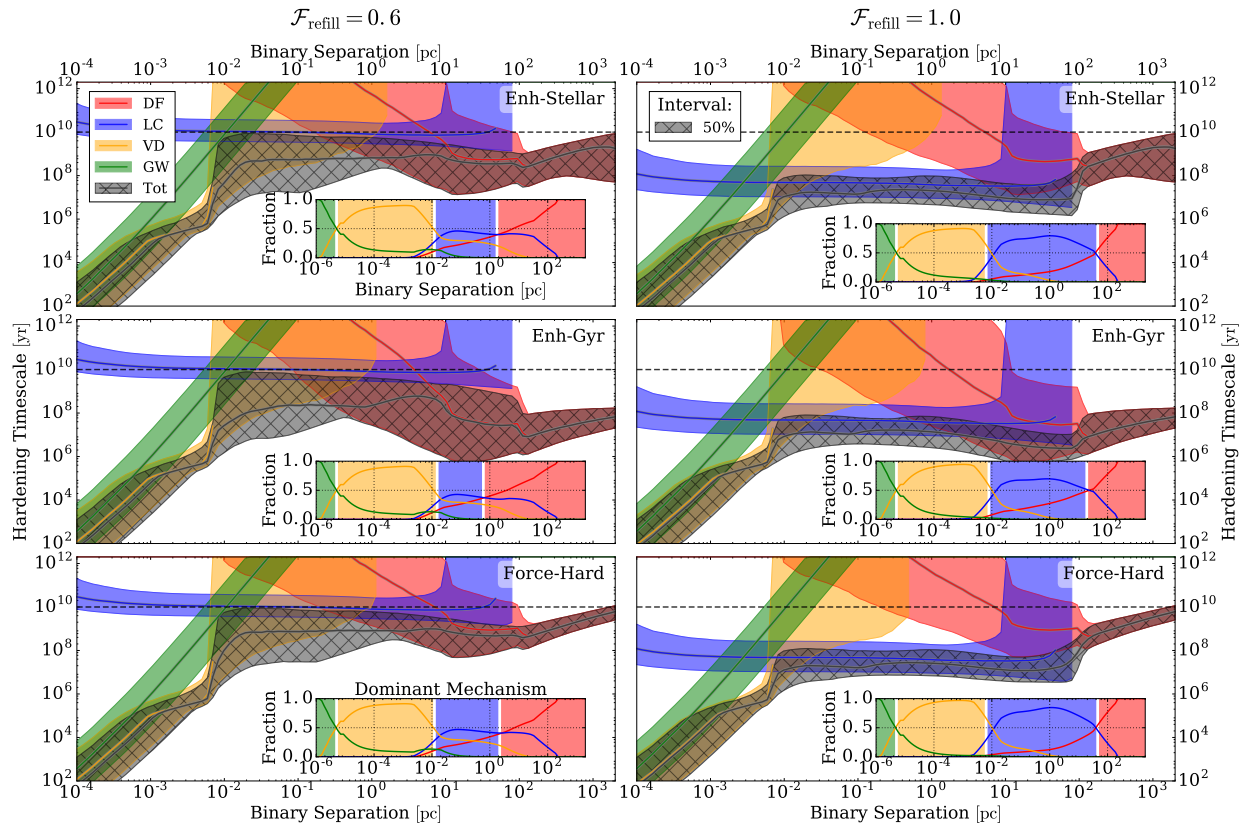
We are grateful to Julie Comerford, Jenny Greene, Daniel Holz, Maura McLaughlin, and Vikram Ravi for fruitful suggestions and discussions. Advice from Alberto Sesana has been extremely helpful throughout, and especially in beginning this project. We are also thankful to the facilitators, organizers, and attendees of the NANOGrav, spring 2016 meeting at CalTech, especially Justin Ellis, Chiara Mingarelli, and Stephen Taylor. We also thank the anonymous referee for numerous, very constructive comments about the manuscript.

This research made use of *Astropy*, a community-developed core Python package for Astronomy ([Astropy Collaboration et al. 2013](#)), in addition to *SciPy* ([Jones et al. 2001](#)), *ipython* ([Prez & Granger 2007](#)), *NumPy* ([van der Walt, Colbert & Varoquaux 2011](#)). All figures were generated using *matplotlib* ([Hunter 2007](#)).

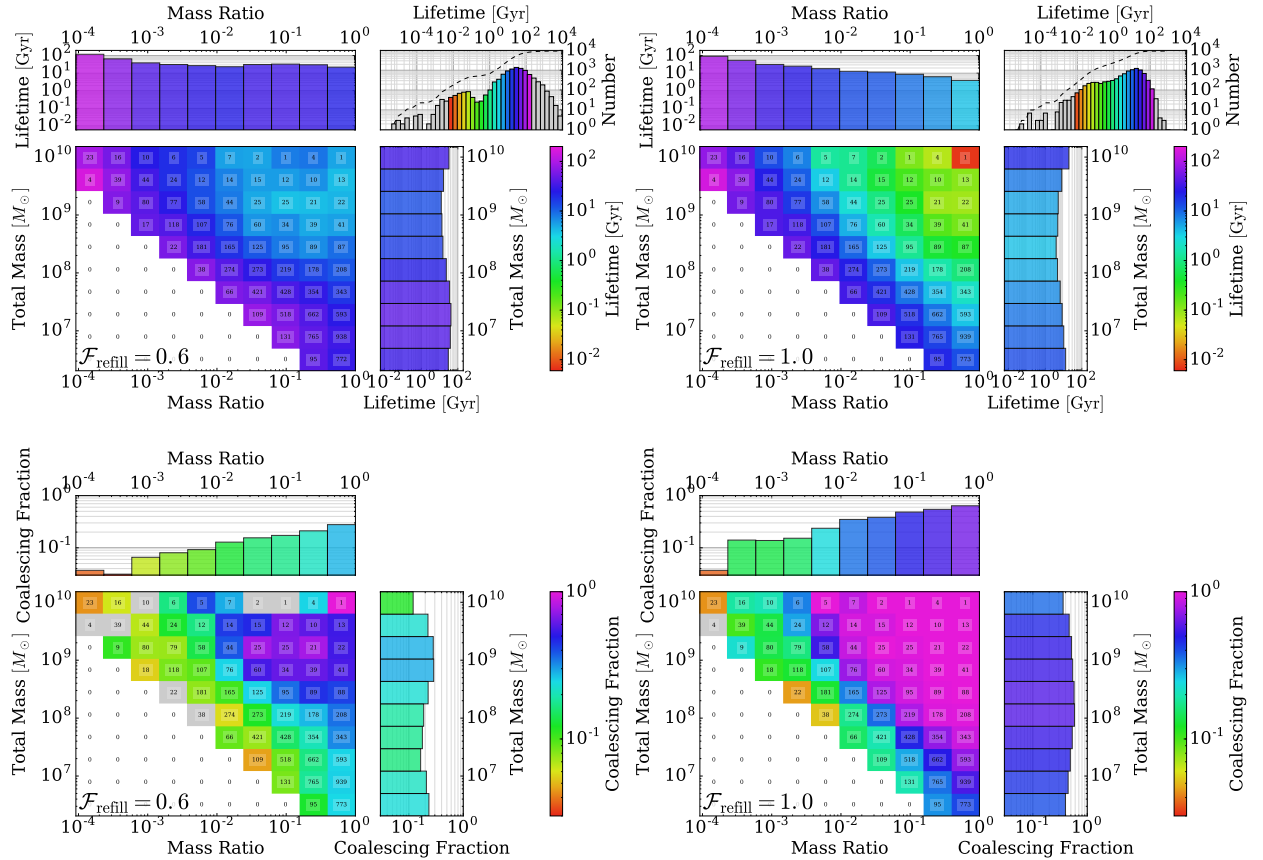


SUPPLEMENTAL MATERIAL

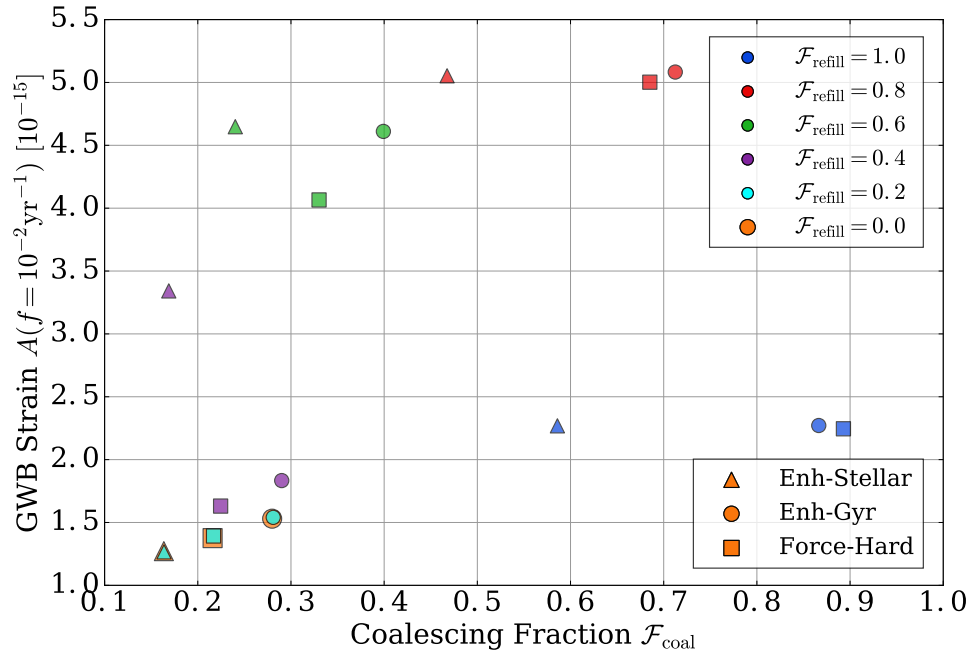
ADDITIONAL FIGURES OF GRAVITATIONAL WAVE SCALINGS AND ALTERNATIVE MODEL RESULTS



**Figure 1.24:** Binary hardening timescales versus separation for different DF models (rows) and LC refilling fractions (columns). Colored lines and bands show the median and 50% intervals for each hardening mechanism: Dynamical Friction (DF), Loss-Cone (LC) scattering, Viscous Drag (VD), and Gravitational Wave (GW) emission with the total hardening rate shown by the grey, hatched region. The inset panels show the fraction of binaries dominated by each mechanism, also as a function of separation. **Note:** in the ‘Force-Hard’ model, the binary separation is artificially altered faster than the timescale shown (between about  $\gtrsim 10^2$  pc). While the DF hardening timescale is still relatively long, the binaries are forcibly hardened at a faster rate.



**Figure 1.25:** Binary lifetimes (upper row) and coalescing fractions (lower row) for our fiducial, moderately refilled loss cone ( $\mathcal{F}_{\text{refill}} = 0.6$ ; left) and an always full one ( $\mathcal{F}_{\text{refill}} = 1.0$ ; right). Both simulations use the ‘Enh-Stellar’ DF model. The overall distribution of MBHB lifetimes are shown in the upper-right-most panel for each simulation, with the cumulative distribution plotted as the dashed line. For  $\mathcal{F}_{\text{refill}} = 0.6$ , most binaries need to be both high total mass ( $M \gtrsim 10^8 M_{\odot}$ ), and moderately high mass ratio ( $\mu \gtrsim 10^{-2}$ ) to have lifetimes short enough to coalesce by redshift zero. In the  $\mathcal{F}_{\text{refill}} = 1.0$  case, on the other hand, either criteria is sufficient—and the coalescing fraction in that parameter space approach unity.



**Figure 1.26:** Dependence of GWB strain amplitude and binary coalescing fraction on LC refilling parameter and DF model. The GW strain is measured at very low frequencies,  $f = 10^{-2} \text{yr}^{-1}$ , and the coalescing fraction is defined using the population of high mass-ratio  $\mu > 0.1$  systems. Each symbol represents a different DF model, and each color a different LC refilling parameter.  $A_{\text{yr}^{-1}}$  tends to increase monotonically with  $\mathcal{F}_{\text{coal}}$ , until the highest  $\mathcal{F}_{\text{refill}}$ . At  $\mathcal{F}_{\text{refill}} \approx 1.0$ , LC stellar scattering is effective enough at these low frequencies to significantly attenuate the GWB amplitude.

SUMMARY OF QUANTITATIVE RESULTS FOR A VARIETY OF PARAMETER CONFIGURATIONS

| Dynamical Friction<br>(Viscous Disk) | Loss Cone                           | GWB   |                       |                     |                                 | Subset                          | Lifetimes [Gyr]<br>Full (Coal) | Coalescing Fraction |               | Persisting Fraction |             |
|--------------------------------------|-------------------------------------|---|-----------------------|---------------------|---------------------------------|---------------------------------|--------------------------------|---------------------|---------------|---------------------|-------------|
|                                      |                                     | Amplitude [ $10^{-16}$ ]<br>$1 \text{ yr}^{-1}$ | $0.1 \text{ yr}^{-1}$ | $1 \text{ yr}^{-1}$ | $0.1 \text{ yr}^{-1}$           |                                 |                                | > 1 pc              | > $10^2$ pc   |                     |             |
| Enh-Stellar<br>(VD: None)            | $\mathcal{F}_{\text{refill}} = 0.6$ | 4.3   | 18                    | -0.64               | -0.55                           | All ( $M > 10^6 M_{\odot}$ )    | 37 (1.3)                       | 1641/9270           | (0.18)        | 0.67                | 0.46        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 32 (1.1)                       | 1089/4759           | (0.23)        | 0.57                | 0.33        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}$            | 26 (1.7)                       | 504/2610            | (0.19)        | 0.68                | 0.45        |
|                                      | $\mathcal{F}_{\text{refill}} = 1.0$ | 6.6   | 25                    | -0.65               | -0.39                           | $M > 10^8 M_{\odot}, \mu > 0.2$ | 11 (0.89)                      | 184/478             | (0.38)        | 0.37                | 0.01        |
|                                      |                                     |   |                       |                     |                                 | All                             | 8.0 (1.3)                      | 4556/9270           | (0.49)        | 0.47                | 0.46        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 5.1 (1.1)                      | 2840/4759           | (0.60)        | 0.35                | 0.33        |
| Enh-Stellar<br>(VD: Fiducial)        | $\mathcal{F}_{\text{refill}} = 0.6$ | <b>3.7</b>                                      | <b>15</b>             | <b>-0.63</b>        | <b>-0.59</b>                    | $M > 10^8 M_{\odot}$            | 5.2 (0.80)                     | 1414/2610           | (0.54)        | 0.45                | 0.45        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}, \mu > 0.2$ | 0.37 (0.35)                    | 469/478             | (0.98)        | 0.01                | 0.01        |
|                                      |                                     |   |                       |                     |                                 | All                             | <b>29 (2.7)</b>                | <b>1875/9270</b>    | <b>(0.20)</b> | <b>0.66</b>         | <b>0.46</b> |
|                                      | $\mathcal{F}_{\text{refill}} = 1.0$ | 4.7   | 17                    | -0.61               | -0.38                           | $\mu > 0.2$                     | <b>26 (2.2)</b>                | <b>1225/4759</b>    | <b>(0.26)</b> | <b>0.55</b>         | <b>0.33</b> |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}$            | <b>17 (4.0)</b>                | <b>608/2610</b>     | <b>(0.23)</b> | <b>0.64</b>         | <b>0.45</b> |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}, \mu > 0.2$ | <b>6.9 (3.3)</b>               | <b>214/478</b>      | <b>(0.45)</b> | <b>0.28</b>         | <b>0.01</b> |
| Enh-Gyr<br>(VD: Fiducial)            | $\mathcal{F}_{\text{refill}} = 0.6$ | 3.8   | 16                    | -0.63               | -0.62                           | All                             | 7.7 (1.2)                      | 4634/9270           | (0.50)        | 0.47                | 0.46        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 4.8 (1.0)                      | 2900/4759           | (0.61)        | 0.35                | 0.33        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}$            | 4.9 (0.79)                     | 1422/2610           | (0.54)        | 0.45                | 0.45        |
|                                      | $\mathcal{F}_{\text{refill}} = 1.0$ | 4.7   | 17                    | -0.61               | -0.38                           | $M > 10^8 M_{\odot}, \mu > 0.2$ | 0.35 (0.35)                    | 472/478             | (0.99)        | 0.01                | 0.01        |
|                                      |                                     |   |                       |                     |                                 | All                             | 13 (1.1)                       | 3536/9270           | (0.38)        | 0.38                | 0.11        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 13 (0.33)                      | 1970/4759           | (0.41)        | 0.30                | 0.06        |
| Force-Hard<br>(VD: Fiducial)         | $\mathcal{F}_{\text{refill}} = 0.6$ | 3.6   | 15                    | -0.64               | -0.61                           | $M > 10^8 M_{\odot}$            | 8.3 (3.5)                      | 1091/2610           | (0.42)        | 0.39                | 0.15        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}, \mu > 0.2$ | 6.4 (3.0)                      | 229/478             | (0.48)        | 0.26                | 0.01        |
|                                      |                                     |   |                       |                     |                                 | All                             | 0.42 (0.30)                    | 7783/9270           | (0.84)        | 0.12                | 0.11        |
|                                      | $\mathcal{F}_{\text{refill}} = 1.0$ | 4.7   | 17                    | -0.61               | -0.38                           | $\mu > 0.2$                     | 0.37 (0.28)                    | 4122/4759           | (0.87)        | 0.07                | 0.06        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}$            | 0.32 (0.25)                    | 2200/2610           | (0.84)        | 0.15                | 0.15        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}, \mu > 0.2$ | 0.20 (0.20)                    | 474/478             | (0.99)        | 0.01                | 0.01        |
| Force-Hard<br>(VD: Fiducial)         | $\mathcal{F}_{\text{refill}} = 0.6$ | 3.6   | 15                    | -0.64               | -0.61                           | All                             | 14 (3.0)                       | 3089/9270           | (0.33)        | 0.32                | 0.02        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 17 (1.9)                       | 1627/4759           | (0.34)        | 0.24                | 0.01        |
|                                      |                                     |   |                       |                     |                                 | $M > 10^8 M_{\odot}$            | 7.8 (4.4)                      | 1093/2610           | (0.42)        | 0.30                | 0.04        |
|                                      | $\mathcal{F}_{\text{refill}} = 1.0$ | 4.7   | 17                    | -0.61               | -0.38                           | $M > 10^8 M_{\odot}, \mu > 0.2$ | 7.0 (3.1)                      | 212/478             | (0.44)        | 0.23                | 0.00        |
|                                      |                                     |   |                       |                     |                                 | All                             | 0.42 (0.30)                    | 7783/9270           | (0.84)        | 0.03                | 0.02        |
|                                      |                                     |   |                       |                     |                                 | $\mu > 0.2$                     | 0.37 (0.28)                    | 4122/4759           | (0.87)        | 0.02                | 0.01        |
| $\mathcal{F}_{\text{refill}} = 1.0$  | 4.7                                 | 17  | -0.61                 | -0.38               | $M > 10^8 M_{\odot}$            | 0.32 (0.25)                     | 2200/2610                      | (0.84)              | 0.05          | 0.04                |             |
|                                      |                                     |   |                       |                     | $M > 10^8 M_{\odot}, \mu > 0.2$ | 0.20 (0.20)                     | 474/478                        | (0.99)              | 0.01          | 0.00                |             |

**Table 1.3:** Summary of quantitative results for the gravitational wave background (GWB), and MBH binary lifetimes & coalescing/persisting fractions. The lifetimes shown are median values of systems in each the ‘Full’ subset, and only those which coalesce by redshift zero (‘Coal’)—the number (and fraction) of such systems are given in the ‘Coalescing Fraction’ column. Finally, the fraction of systems which remain uncoalesced at separations  $r > 1$  pc and  $r > 10^2$  pc are shown in the ‘Persisting Fraction’ column. Results for a simulation using models with all fiducial parameters are shown in bold.

## STELLAR, LOSS-CONE (LC) SCATTERING CALCULATIONS

The rate at which stars can refill the loss cone is governed by the ‘relaxation time’ ( $\tau_{\text{rel}}$ ). Following [Binney & Tremaine \(1987\)](#), consider a system of  $N$  masses  $m$ , with number density  $n$ , and characteristic velocities  $v$ . The relaxation time can be written as,

$$\tau_{\text{rel}} \approx \frac{N}{8 \ln \Lambda} \tau_{\text{cross}} \approx \frac{v^3}{8\pi G^2 m^2 n \ln \Lambda}, \quad (1.19)$$

where  $\ln \Lambda$  is again the Coulomb Logarithm, and  $\tau_{\text{cross}} \equiv r/v$  is the crossing-time.  $\tau_{\text{rel}}$  represents the characteristic time required to randomize a particle’s velocity via scatterings, i.e., Eq. 1.19 can be used to define the diffusion coefficient  $\mathcal{D}_{v^2}$  as,  $\tau_{\text{rel}} \approx v^2/\mathcal{D}_{v^2}$ . If  $t/\tau_{\text{rel}} \ll 1$ , then two-body encounters (and relaxation) haven’t been important.

Consider a distribution function (or phase-space density)  $f = f(\vec{x}, \vec{v})$ , such that the number of stars in a small spatial-volume  $d^3\vec{x}$  and velocity-space volume  $d^3\vec{v}$  is given as  $f(\vec{x}, \vec{v}) d^3\vec{x} d^3\vec{v}$ . In a spherical system in which there are the conserved energy  $E$  and angular momentum  $\vec{L}$ , the six independent position and velocity variables can be reduced to these four independent, conserved quantities via the Jeans theorem. Furthermore, if the system is perfectly spherically symmetric—which we assume in our analysis, then the three independent angular momentum components can be replaced with the angular momentum magnitude, i.e.  $f = f(E, L)$ .

If we define a relative potential and relative energy,  $\Psi \equiv -\Phi + \Phi_0$ , and,  $\mathcal{E} \equiv -E + \Phi_0 = \Psi - \frac{1}{2}v^2$ , then we can calculate the number density as\*,

$$\begin{aligned} n(\vec{x}) = n(x) &= 4\pi \int_0^{\sqrt{2\Psi}} f(x, v) v^2 dv \\ &= 4\pi \int_0^{\Psi} f(\mathcal{E}) [2(\Psi - \mathcal{E})]^{1/2} d\mathcal{E}. \end{aligned} \quad (1.20)$$

Inverting this relationship, the distribution function can be calculated from an isotropic density profile using,

$$\begin{aligned} f(\mathcal{E}) &= \frac{1}{\pi^2 \sqrt{8}} \frac{d}{d\mathcal{E}} \int_0^{\mathcal{E}} \frac{dn}{d\Psi} \frac{d\Psi}{(\mathcal{E} - \Psi)^{1/2}} \\ &= \frac{1}{\pi^2 \sqrt{8}} \left[ \mathcal{E}^{-1/2} \left( \frac{dn}{d\Psi} \right)_{\Psi=0} + \int_0^{\mathcal{E}} \frac{d^2n}{d\Psi^2} \frac{d\Psi}{(\mathcal{E} - \Psi)^{1/2}} \right]. \end{aligned} \quad (1.21)$$

We have found the latter form of (1.21) to be much simpler and more reliable to implement.

We follow the discussion and prescription for loss-cone scattering given by [Magorrian & Tremaine \(1999\)](#), corresponding to a single central object in a spherical (isotropic) background of stars. We adapt this prescription simply by modifying the radius of interaction to be appropriate for scattering with a binary instead of being tidally disrupted by a single MBH. A more extensive discussion of loss-cone dynamics—explicitly considering MBH binary systems and asphericity—can be found in [Merritt \(2013\)](#).

Stars with a pericenter distance smaller than some critical radius  $\mathcal{R}_{\text{crit}}$  will interact with the binary. For a fixed energy ( $\mathcal{E}$ ) orbit, there is then a critical angular momentum,  $J_{lc}(\mathcal{E}) = \mathcal{R}_{\text{crit}} (2[\mathcal{E} - \Psi(\mathcal{R}_{\text{crit}})])^{1/2} \approx \mathcal{R}_{\text{crit}} (2[-\Psi(\mathcal{R}_{\text{crit}})])^{1/2}$ , which defines the loss-cone ([Frank & Rees 1976](#); [Lightman & Shapiro 1977](#)). In general, the number of stars with energy and angular momentum in the range  $d\mathcal{E}$  and  $dJ^2$  around  $\mathcal{E}$  and  $J^2$  can be calculated as,

$$N(\mathcal{E}, J^2) d\mathcal{E} dJ^2 = 4\pi^2 f(\mathcal{E}, J^2) \cdot P(\mathcal{E}, J^2) d\mathcal{E} dJ^2, \quad (1.22)$$

where  $P(\mathcal{E}, J^2)$  is the stellar orbital period. For an isotropic stellar distribution  $f(\mathcal{E}, J^2) = f(\mathcal{E})$ , and

\*  $\Phi_0$  is arbitrary, but  $\Phi_0 \equiv E(r \rightarrow \infty)$  may be convenient.

$P(\mathcal{E}, J^2) \approx P(\mathcal{E})$ . The total number of stars can be calculated as,

$$N_i(\mathcal{E}) d\mathcal{E} = 4\pi^2 f(\mathcal{E}) P(\mathcal{E}) J_i^2(\mathcal{E}) d\mathcal{E}. \quad (1.23)$$

For the number of stars in the loss-cone  $N_{lc}(\mathcal{E})$ , this uses the LC angular momentum,  $J_i^2(\mathcal{E}) = J_{lc}^2(\mathcal{E})$ ; and for all stars  $N(\mathcal{E})$ , the circular (and thus maximum) angular momentum,  $J_i^2(\mathcal{E}) = J_c^2(\mathcal{E})$ . When we initially calculate the distribution function, we use the stellar density profile from Illustris galaxies which have recently hosted a MBH ‘merger’ event (see: §1.2.2). We assume that the resulting distribution function  $f(\mathcal{E})$  is (so-far) unperturbed by the MBH binary, i.e. it does not take into account stars already lost (scattered). The resulting  $N_{lc}(\mathcal{E})$  from Eq. 1.23 then corresponds to the number of stars in the ‘full’ loss-cone specifically.

Stars in the LC are consumed on their orbital timescale  $\tau_{orb} = P(\mathcal{E})$ . The rate of flux of stars to within  $\mathcal{R}_{crit}$  is then,

$$F_{lc}^{full}(\mathcal{E}) d\mathcal{E} = 4\pi^2 f(\mathcal{E}) J_{lc}^2(\mathcal{E}) d\mathcal{E}, \quad (1.24)$$

coming almost entirely from within the central objects sphere of influence  $\mathcal{R}_{infl}$ , defined as  $M(r < \mathcal{R}_{infl}) \approx M_\bullet$ . Refilling of the loss-cone occurs on the characteristic relaxation timescale  $\tau_{rel}$ . From Eq. 1.19, it is clear that  $\tau_{orb}/\tau_{rel} \approx \tau_{cross}/\tau_{rel} \ll 1$ , i.e. the loss-cone is drained significantly faster than it is refilled—and the loss-cone will, in general, be far from ‘full’.

To calculate the steady-state flux of the loss-cone, the Fokker-Planck equation must be solved with a fixed (unperturbed) background stellar distribution at the outer edge of the LC and no stars surviving within the scattering region at the inner-edge. A full derivation can be found in Magorrian & Tremaine (1999), which yields a equilibrium flux of stars,

$$F_{lc}^{eq}(\mathcal{E}) d\mathcal{E} = 4\pi^2 P(\mathcal{E}) J_c^2(\mathcal{E}) f(\mathcal{E}) \frac{\mu(\mathcal{E})}{\ln R_0^{-1}(\mathcal{E})}, \quad (1.25)$$

where the angular momentum diffusion parameter  $\mu \equiv 2r^2 \mathcal{D}_{v^2} / J_c^2$ , and,

$$\ln R_0^{-1} = -\ln \mathcal{R}_{crit} + \begin{cases} q & q \geq 1 \\ 0.186q + 0.824\sqrt{q} & q < 1 \end{cases}, \quad (1.26)$$

describes the effective refilling radius depending on which refilling regime (‘pin-hole’ or ‘diffusive’, see Fig. 1 of Lightman & Shapiro 1977) is relevant, for a refilling parameter  $q(\mathcal{E}) \equiv P(\mathcal{E}) \mu(\mathcal{E}) / \mathcal{R}_{crit}(\mathcal{E})$ .

Equations 1.24 & 1.25 give the full and steady-state LC fluxes, which are interpolated between using a logarithmic, ‘refilling fraction’ (Eq. 1.11) which then determines the hardening rate of each binary in our simulations.



# THE GRAVITATIONAL WAVE BACKGROUND FROM MASSIVE BLACK HOLE BINARIES IN ILLUSTRIS: SPECTRAL FEATURES AND TIME TO DETECTION WITH PULSAR TIMING ARRAYS

*This thesis chapter originally appeared in the literature as*

Luke Zoltan Kelley, Laura Blecha, Lars Hernquist, Alberto Sesana, Stephen R. Taylor  
MNRAS, 471, 4. arXiv:1702.02180

## ABSTRACT

PULSAR Timing Arrays (PTA) around the world are using the incredible consistency of millisecond pulsars to measure low frequency gravitational waves from (super)Massive Black Hole (MBH) binaries. We use comprehensive MBH merger models based on cosmological hydrodynamic simulations to predict the spectrum of the stochastic Gravitational-Wave Background (GWB). We use real Time-of-Arrival (TOA) specifications from the European, NANOGrav, Parkes, and International PTA (IPTA) to calculate realistic times to detection of the GWB across a wide range of model parameters. In addition to exploring the parameter space of environmental hardening processes (in particular: stellar scattering efficiencies), we have expanded our models to include eccentric binary evolution which can have a strong effect on the GWB spectrum. Our models show that strong stellar scattering and high characteristic eccentricities enhance the GWB strain amplitude near the PTA sensitive “sweet-spot” (near the frequency  $f = 1 \text{ yr}^{-1}$ ), slightly improving detection prospects in these cases. While the GWB *amplitude* is degenerate between cosmological and environmental parameters, the location of a spectral turnover at low frequencies ( $f \lesssim 0.1 \text{ yr}^{-1}$ ) is strongly indicative of environmental coupling. At high frequencies ( $f \gtrsim 1 \text{ yr}^{-1}$ ), the GWB spectral index can be used to infer the number density of sources and possibly their eccentricity distribution. Even with merger models that use pessimistic environmental and eccentricity parameters, if the current rate of PTA expansion continues, we find that the International PTA is highly likely to make a detection within about 10 years.

## 2.1 INTRODUCTION

Pulsar Timing Arrays (PTA) are expected to detect Gravitational Waves (GW; [Sazhin 1978](#); [Dewtweiler 1979](#); [Romani & Taylor 1983](#)) from stable binaries of (super)-Massive Black Holes (MBH; [Rajagopal & Romani 1995](#); [Wyithe & Loeb 2003](#); [Phinney 2001](#)). These arrays use correlated signals in the consistently timed pulses from millisecond pulsars to search for low-frequency ( $\lesssim 10 \text{ yr}^{-1}$ ) perturbations to flat space-time ([Hellings & Downs 1983](#); [Foster & Backer 1990](#)). There are currently three independent PTA searching for GW signals: the North-American Nanohertz Observatory for Gravitational waves (NANOGrav; [McLaughlin 2013](#)), the European PTA (EPTA; [Kramer & Champion 2013](#)), and the Parkes PTA (PPTA; [Manchester et al. 2013](#)). Additionally, the International PTA (IPTA, [Hobbs et al. 2010](#)) is a collaboration which aims to combine the data and expertise from each independent group.

Comparable upper limits on the presence of a stochastic Gravitational Wave Background (GWB) have been calculated by the EPTA ([Lentati et al. 2015](#)), NANOGrav ([Arzoumanian et al. 2016](#)), PPTA ([Shannon et al. 2015](#)) and IPTA ([Verbiest et al. 2016](#)). These upper limits are already astrophysically informative in that much of the previously predicted parameter space is now in tension with observations, and there are suggestions that some models are excluded ([Shannon et al. 2015](#)). Many previous GWB models have assumed that most or all of the MBH pairs formed after the merger of their host galaxies are able to quickly reach the ‘hard binary’ phase\* ( $\lesssim 10 \text{ pc}$ ) and eventually coalesce due to GW emission (e.g. [Wyithe & Loeb 2003](#); [Jaffe & Backer 2003](#); [Sesana 2013b](#)). These models, which assume GW-only driven evolution and produce purely power-law GWB spectra, likely over-predict the GWB energy in that: 1) perhaps a substantial fraction of MBH Binaries stall at galactic scales ( $\sim \text{kpc}$ ), or before reaching the small separations ( $\sim 10^{-3}\text{--}10^{-1} \text{ pc}$ ) corresponding to the PTA sensitive band (e.g. [McWilliams, Ostriker & Pretorius 2014](#)); and 2) significant ‘attenuation’ of the GW signal may exist due to environmental processes (non-GW hardening, due to stellar scattering or coupling with a circumbinary gaseous disk) which decrease the amount of time binaries spend in a given frequency interval (e.g. [Kocsis & Sesana 2011](#); [Sesana 2013b](#); [Ravi et al. 2014](#); [Rasskazov & Merritt 2016](#)).

Some recent GW-only models predict lower signal levels because of differing cosmological assumptions (i.e. galaxy-galaxy merger rates, the mass functions of MBH, etc) which produce different distributions of MBH binaries (e.g. [Roebber et al. 2016](#); [Sesana et al. 2016](#)), eliminating the tension with PTA upper limits. More comprehensive models have also been assembled which take into account binary-stalling and GW-attenuation (e.g. [Ravi et al. 2014](#)). Some of these models suggest that the GWB is only just below current observational sensitivities, which begs the question, ‘*how long until we make a detection?*’ Recently, PTA detection statistics conveniently formalized in [Rosado, Sesana & Gair \(2015\)](#) have been used by [Taylor et al. \(2016b\)](#) to calculate times to detections for purely power-law, GW-only GWB models with a full range of plausible GWB amplitudes. [Vigeland & Siemens \(2016\)](#) also calculate detection statistics using a more extensive suite of broken power-laws to model the effects of varying environmental influences.

In [Kelley, Blecha & Hernquist \(2017, hereafter ‘Paper-1’\)](#) we construct the most comprehensive MBHB merger models to date, using the self-consistently derived population of galaxies and MBH from the Illustris cosmological, hydrodynamic simulations (§2.2.1, e.g. [Vogelsberger et al. 2014a](#); [Genel et al. 2014](#)). The MBHB population is post-processed using semi-analytic models of GW emission in addition to environmental hardening mechanisms that are generally *required* for MBHB to reach small separations within a Hubble time (e.g. [Begelman, Blandford & Rees 1980](#); [Milosavljević & Merritt 2003](#)), and emit GW in PTA-sensitive frequency bands.

\*‘Hard’ binaries are distinguished by, and important because, scattering interactions tend to further harden the binary (e.g. [Hut 1983](#)).

In this paper, we introduce the addition of eccentric binary evolution to our models, and explore its effects on the GWB. To produce more realistic GWB spectra, we complement our previous semi-analytic (SA) calculations with a more realistic, Monte-Carlo (MC) technique. Using our merger models and resulting spectra, we calculate realistic times to detection for each PTA following [Rosado, Sesana & Gair \(2015\)](#) & [Taylor et al. \(2016b\)](#). In §2.2 we describe our MBHB, GWB and PTA models. Then in §2.3 we describe the effects of eccentricity on binary evolution (§2.3.1) and the GWB spectrum (§2.3.2) including comparisons between the SA and MC calculations, and finally our predictions for times to GWB detections (§2.3.3).

## 2.2 METHODS

Our simulations use the coevolved galaxies and MBH particles from the Illustris cosmological, hydrodynamic simulations ([Vogelsberger et al. 2013](#); [Torrey et al. 2014](#); [Vogelsberger et al. 2014b](#); [Genel et al. 2014](#); [Sijacki et al. 2015](#)) run using the Arepo ‘moving-mesh’ code ([Springel 2010](#)). Our general procedure of extracting MBH, their merger events and their galactic environments are described in detail in [Paper-1](#). Here, we give a brief overview of our methods (§2.2.1) and the improvements made to include eccentric binary evolution (§2.2.2). We then describe the methods by which we calculate GW signatures (§2.2.3) and realistic detection statistics for simulated PTA (§2.2.4).

### 2.2.1 ILLUSTRIS MBH MERGERS AND ENVIRONMENTS

The Illustris simulation is a cosmological box of 106.5 Mpc on a side (at  $z = 0.0$ ) containing moving-mesh gas cells, and particles representing stars, dark matter, and MBH. All of the Illustris data is publicly available online ([Nelson et al. 2015](#)). MBH are ‘seeded’ with a mass of  $1.42 \times 10^5 M_{\odot}$  into halos with masses above  $7.1 \times 10^{10} M_{\odot}$  ([Sijacki et al. 2015](#)), where they accrete gas from the local environment and grow over time. As they develop, they proportionally deposit energy back into the local environment ([Vogelsberger et al. 2013](#)). When two MBH particles come within a gravitational smoothing length of one another (typically on the order of a kpc) a ‘merger’ event is recorded. From those mergers we identify the constituent MBH and the host galaxy in which they subsequently reside. From the host galaxy, density profiles are constructed which are used to determine the environment’s influence on the MBHB merger process. The simulations used here, as in [Paper-1](#), are semi-analytic models which integrate each binary (independently) from large-scale separations down to eventual coalescence based on prescriptions for GW- and environmentally- driven hardening.

### 2.2.2 MODELS FOR ECCENTRIC BINARY EVOLUTION

We implement four distinct mechanisms which dissipate orbital energy and ‘harden’ the MBHB (as in [Paper-1](#)):

- **Dynamical Friction (DF)**, dominant on  $\sim$  kpc scales) is implemented following [Chandrasekhar \(1943\)](#) and [Binney & Tremaine \(1987\)](#), based on the local density (gas and dark matter) and velocity dispersion. The mass of the decelerating object is taken as the mass of the secondary MBH along with its host galaxy. We assume a model for tidal stripping such that the decelerating mass decreases as a power-law from the combined mass, to that of only the secondary MBH, over the

course of a dynamical time\*. With the addition of eccentric binary evolution, we make the approximation that DF does not noticeably affect the eccentricity distribution of binaries (e.g. Colpi, Mayer & Governato 1999; van den Bosch et al. 1999; Hashimoto, Funato & Makino 2003), and that the semi-major axis remains the relevant distance scale. Equivalently, the “initial” eccentricities in our models can be viewed as the eccentricity once binaries enter the stellar scattering regime.

- **Stellar Loss-Cone (LC) scattering** ( $\sim$  pc), is implemented using the prescription from Sesana, Haardt & Madau (2006) & Sesana (2010) following the formalism of Quinlan (1996). Here, the hardening rate ( $da/dt$ ) and eccentricity evolution ( $de/dt$ ) are determined by dimensionless constants  $H$  and  $K$ , calculated in numerical scattering experiments such that,

$$\left. \frac{da}{dt} \right|_u \equiv -\frac{G\rho}{\sigma} a^2 H, \quad (2.1)$$

and

$$\left. \frac{de}{dt} \right|_u \equiv \frac{G\rho}{\sigma} a H K. \quad (2.2)$$

Here the binary separation (semi-major axis,  $a$ ) and eccentricity ( $e$ ) are evolved based on profiles of density ( $\rho$ ) and velocity dispersion ( $\sigma$ ) calculated from each binary host-galaxy in Illustris. We use the fitting formulae and tabulated constants for  $H$  and  $K$  from Sesana, Haardt & Madau (2006)<sup>†</sup>. Note that this semi-empirical approach, which we will refer to as ‘*eccentric LC models*’, explicitly assumes a full loss-cone in the scattering experiments by which they are calibrated.

In our previous calculations presented in Paper-1, we used a different LC prescription for binaries restricted to circular orbits. In this paper we focus on the *eccentric LC models*, but include results with our previous ‘*circular*’ prescription for comparison. The circular models<sup>‡</sup> include a dimensionless ‘refilling parameter’,  $\mathcal{F}_{\text{refill}} \in \{0.0, 1.0\}$ , which interpolates between a ‘steady-state’ LC ( $\mathcal{F}_{\text{refill}} = 0.0$ ), where equilibrium is reached between the scattering rate and refilling by the two-body diffusion of stars; and a ‘full’ LC ( $\mathcal{F}_{\text{refill}} = 1.0$ ), where the stellar distribution function is unaltered by the presence of the scattering source. The  $\mathcal{F}_{\text{refill}}$  parameter for true astrophysical binaries is highly uncertain, but can drastically affect the efficiency with which binaries coalesce (see the discussion in Paper-1). We explore six models with,  $\mathcal{F}_{\text{refill}} = [0.0, 0.2, 0.4, 0.6, 0.8, 1.0]$ . Some recent studies tend to favor nearly-full LC models (e.g. Sesana & Khan 2015; Vasiliev, Antonini & Merritt 2015). The eccentric LC model, with imposed zero-eccentricities ( $e_0 = 0.0$ ), yields results very similar to the circular-only model with a full LC ( $\mathcal{F}_{\text{refill}} = 1.0$ ), as expected.

- **Gas drag from a circumbinary, Viscous Disk (VD;  $\sim 10^{-3}$  pc)** is calculated following the thin disk models from Haiman, Kocsis & Menou (2009). In these models, the disk is composed of three, physically distinct regions (Shapiro & Teukolsky 1986) determined by the dominant pressure (radiation versus thermal) and opacity (Thomson versus free-free) sources. From inner- to outer-disk, the regions are: 1) radiation & Thomson, 2) thermal & Thomson, and 3) thermal & free-free.

\*The dynamical time used is that of the primary’s host galaxy. This corresponds to the ‘Enh-Stellar’ model from Paper-1.

<sup>†</sup>Sesana, Haardt & Madau (2006): Eqs. 16 & 18, and Tables 1 & 3, respectively

<sup>‡</sup>These follow the theoretically-derived formulae from Magorrian & Tremaine (1999) for a spherically-symmetric background of stars scattering with a central object. Scattering rates are calculated assuming isotropic, Maxwellian velocities and stellar distribution functions calculated from each galaxy’s stellar density profile. The density profiles are extended to unresolved ( $\lesssim$  pc) scales with power-law extrapolations. Additional details on Illustris stellar densities are included in §2.4.

The disk density profiles are constructed based on the self-consistently derived accretion rates given by Illustris, and truncated based on a (Toomre) gravitational stability criterion. Higher densities resulting from higher accretion rates lead to more extended inner-disk regions. We use an alpha-disk\* throughout. It's worth noting that the inner-disk region (1) has a very similar hardening curve to GW-emission:  $\tau_{\text{VD},1} \propto r^{7/2}$ , versus  $\tau_{\text{gw}} \propto r^4$ . In Illustris, post-merger MBH tend to have higher accretion rates, and thus larger inner-disk regions.

We assume that the disk has a negligible effect on the eccentric evolution of binaries, i.e.  $[de/dt]_{\text{VD}} = 0$ . This assumption is made for simplicity. While numerous studies have shown that eccentric evolution can at times be significant in circumbinary disks (e.g. [Armitage & Natarajan 2005](#); [Cuadra et al. 2009a](#); [Roedig et al. 2011](#)), we are unaware of generalized descriptions of eccentricity evolution for binaries/disks with arbitrary initial configurations. In the analysis which follows, we explore a wide range of eccentricity parameter space. While a given model may end up being inconsistent with VD eccentric-evolution, the overall parameter space should still encompass the same resulting GWB spectra.

- **Gravitational Wave (GW) emission** ( $\sim 10^{-5}$  pc) hardens binaries at a rate given by [Peters \(1964, Eq. 5.6\)](#) as,

$$\frac{da}{dt} = -\frac{64 G^3 M_1 M_2 (M_1 + M_2)}{5 c^5 a^3} F(e), \quad (2.3)$$

where the eccentric enhancement,

$$F(e) \equiv \frac{(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4)}{(1 - e^2)^{7/2}}. \quad (2.4)$$

In [Paper-1](#) we made the approximation that the eccentricity of all binaries was negligible and thus  $F(e) = 1$ . Here, we include models with non-zero eccentricity, evolved as ([Peters 1964, Eq. 5.7](#)),

$$\frac{de}{dt} = -\frac{304 G^3 M_1 M_2 (M_1 + M_2)}{15 c^5 a^4} \frac{(e + \frac{121}{304}e^3)}{(1 - e^2)^{5/2}}. \quad (2.5)$$

### 2.2.3 GRAVITATIONAL WAVES FROM ECCENTRIC MBH BINARIES

Circular binaries, with (rest-frame) orbital frequencies  $f_r$ , emit GW monochromatically at  $2f_r$ , i.e. the  $n = 2$  harmonic. Eccentric systems lose symmetry, and emit at  $n = 1$  and all higher harmonics, i.e.  $f_h = n f_r$  (for  $n \in \mathbb{I}$ ). The GW energy spectrum can then be expressed as ([Enoki & Nagashima 2007, Eq. 3.10](#)),

$$\frac{d\varepsilon_{\text{GW}}}{df_r} = \sum_{n=1}^{\infty} \left[ L_{\text{GW,circ}}(f_h) \frac{\tau_h^f(f_h, e)}{n f_h} g(n, e) \right]_{f_h=f_r/n}. \quad (2.6)$$

The GW frequency-distribution function  $g(n, e)$  is shown in [Eq. 2.19](#). Equation (2.6) describes the GW spectrum emitted by a binary *over its lifetime*, which is used in the semi-analytic GWB calculation (§2.2.3.1). The total power radiated by an eccentric binary is enhanced by the factor  $F(e)$ , i.e.,  $L_{\text{GW}}(f_r, e) = L_{\text{GW,circ}}(f_r) \cdot F(e)$ , where the GW luminosity for a circular binary is ([Peters & Mathews 1963, Eq. 16](#)),

$$L_{\text{GW,circ}}(f_r) = \frac{32}{5 G c^5} (G M 2\pi f_r)^{10/3}. \quad (2.7)$$

\*Where viscosity depends on both gas and radiation pressure, as apposed to a ‘beta-disk’ which depends only on the gas-pressure. The differences in merger times and coalescing fractions between the alpha and beta models are negligible. We use the alpha model because it may be more conservative via higher viscosities in the inner-most disk regions which, while insignificant for increasing the number of merging MBH, could increase GWB attenuation.

Note that in Eq. 2.6, the relevant timescale is the hardening-time (or ‘residence’-time) *in frequency*,

$$\tau_h^f \equiv \left| \frac{f}{df/dt} \right| = \frac{2}{3} \left| \frac{a}{da/dt} \right| \equiv \frac{2}{3} \tau_h, \quad (2.8)$$

which is 2/3 the hardening-time *in separation* (via Kepler’s law), which we use for most of our discussion and figures.

The GW strain from an individual, eccentric source can be related to that of a circular source as (e.g. Amaro-Seoane et al. 2010, Eq. 9)\*,

$$h_s^2(f_r) = \sum_{n=1}^{\infty} h_{s,\text{circ}}^2(f_h) \left( \frac{2}{n} \right)^2 g(n, e) \Big|_{f_h=f_r/n}. \quad (2.9)$$

Here, the GW strain from a circular binary is,

$$h_{s,\text{circ}}(f_r) = \frac{8}{10^{1/2}} \frac{(GM)^{5/3}}{c^4 d_L} (2\pi f_r)^{2/3} \quad (2.10)$$

(e.g. Sesana, Vecchio & Colacino 2008, Eq. 8), for a luminosity distance  $d_L$ , and a chirp mass  $\mathcal{M} = (M_1 M_2)^{3/5} / (M_1 + M_2)$ . Equation (2.9) describes the *instantaneous* GW strain amplitude from a binary, and is used in the Monte-Carlo GWB calculation (§2.2.3.2).

The GWB is usually calculated in one of two ways (Sesana, Vecchio & Colacino 2008): either Semi-Analytically (SA), treating the distribution of binaries as a smooth, continuous and deterministic function to calculate  $\partial^5 n_c(M_1, M_2, z, f_r, e) / \partial M_1 \partial M_2 \partial z \partial f_r \partial e$  (Phinney 2001), or alternatively, in the Monte Carlo (MC) approach, where  $n_c(M_1, M_2, z, f_r, e)$  is considered as a particular realization of a finite number of MBHB in the universe (Rajagopal & Romani 1995).

### 2.2.3.1 SEMI ANALYTIC GWB

The GWB spectrum can be calculated from a distribution of eccentric binaries as (Huerta et al. 2015; Enoki & Nagashima 2007, Eq. 3.11),

$$h_c^2(f) = \frac{4G}{\pi c^2 f} \int dM_1 dM_2 dz n_c(M_1, M_2, z) \sum_{n=1}^{\infty} \left[ L_{\text{GW,circ}}(f_r) \frac{\tau_h^f}{n f_r} g(n, e) \right]_{f_r=f(1+z)/n}, \quad (2.11)$$

where the summation is evaluated for all rest-frame frequencies with a harmonic matching the observed (redshifted) frequency bin  $f$ . Eq. 2.11 is derived by integrating the emission of each binary over its lifetime, which is assumed to happen quickly ( $\tau_h^f \ll \tau_{\text{Hubble}}$ ).

Each of the binaries in our simulation is evolved from their formation time (identified in Illustris) until coalescence. The GWB calculations only include the portions of the evolution which occur before redshift zero. In our implementation of the Eq. 2.11 calculation, interpolants are constructed for each binary’s parameters (e.g. frequency, GW strain, etc) over its lifetime which are then used when sampling by simulated PTA.

### 2.2.3.2 MONTE CARLO GWB

The GWB spectrum can also be constructed as the sum of individual source strains for all binaries emitting at the appropriate frequencies (and harmonics) in the observer’s past light cone (Sesana, Vecchio

\*Note the factor of  $(2/n)^2$  when converting from circular to eccentric systems.

& Colacino 2008, Eq. 6 & 10),

$$\begin{aligned} h_c^2(f) &= \int dz d\mathcal{M} \left[ \frac{d^3 N}{dz d\mathcal{M} d \ln f_r} h_s^2(f_r) \right]_{f_r=f(1+z)} \\ &= \int dz d\mathcal{M} \frac{d^2 n_c}{dz d\mathcal{M}} dV_c \left[ \frac{f_r}{df_r} h_s^2(f_r) \right]_{f_r=f(1+z)}, \end{aligned} \quad (2.12)$$

for a number of sources  $N$ , or comoving number-density  $n_c$  in a comoving volume  $V_c$ .

The differential element of the past light cone can be expressed as (e.g. Hogg 1999, Eq. 28),

$$dV_c(z) = 4\pi (1+z)^2 \frac{c}{H_0} \frac{d_c^2(z)}{E(z)} dz, \quad (2.13)$$

for a comoving distance  $d_c = d_L/(1+z)$ , redshift-zero Hubble constant  $H_0$ , and the cosmological evolution function  $E(z)$  (given in Eq. 2.21). The term  $f_r/df_r = f/df$ , which results from the *definition* of the characteristic strain as that over a logarithmic frequency interval, can be identified as the number of cycles each binary spends emitting in a given frequency interval (see Eq. 2.20).

To discretize Eq. 2.12 for a quantized number of sources (e.g. from a simulation) we convert the integral over number density, into a sum over sources within the Illustris comoving volume  $V_{\text{ill}}$ ,

$$\int dz d\mathcal{M} \dots \frac{d^2 n_c}{dz d\mathcal{M}} dV_c \rightarrow \sum_{ij} \dots \frac{\Delta V_{ij}}{V_{\text{ill}}}, \quad (2.14)$$

where the summation is over all binaries  $i$  at each time-step  $j$ . The factor  $\frac{\Delta V_{ij}}{V_{\text{ill}}} \equiv \Lambda_{ij}$  represents *the number of MBH binaries in the past light cone represented by each binary in the simulation\**. The volume of the past light cone represented by  $\Lambda_{ij}$  depends on the integration step-size, i.e.,

$$\Lambda_{ij} = \frac{1}{V_{\text{ill}}} \frac{dV_c(z_{ij})}{dz_{ij}} \Delta z_{ij}, \quad (2.15)$$

where  $\Delta z_{ij}$  is the redshift step-size for binary  $i$  at time step  $j$ .  $\Lambda_{ij}$  is stochastic, determined by the number of binaries in a given region of the universe. Alternative ‘realizations’ of the universe can be constructed by, instead of using  $\Lambda_{ij}$  itself, scaling by a factor drawn from a Poisson distribution  $\mathcal{P}$ , centered at  $\Lambda_{ij}$ .

Thus, to construct a particular realization of the GWB spectrum we calculate,

$$h_c^2(f) = \sum_{ij} \mathcal{P}(\Lambda_{ij}) \sum_{n=1}^{\infty} \left[ \frac{f_r}{\Delta f} h_s^2(f_r) \left( \frac{2}{n} \right)^2 g(n, e) \right]_{f_r=f(1+z)/n}. \quad (2.16)$$

## 2.2.4 DETECTION WITH PULSAR TIMING ARRAYS

To calculate the detectability of our predicted GWB spectra, we use the detection formalism outlined by Rosado, Sesana & Gair (2015). A ‘detection statistic’<sup>†</sup>  $X$  is constructed as the cross-correlation of PTA data using a filter which maximizes the Detection Probability (DP)  $\gamma$ . The optimal filter is known to be the ‘overlap reduction function’ (Finn, Larson & Romano 2009) which, for PTA, is the Hellings & Downs (1983) curve that depends on the particular PTA configuration (angular separation between each pair of pulsars). Using the optimal detection statistic, and the noise characteristics of the PTA under consideration, parameters like the signal-to-noise ratio (SNR; and SNR-threshold) or DP can be calculated based on a GWB. Rosado, Sesana & Gair (2015) should be consulted for the details of the detection formalism but, for completeness, the relevant equations used in our calculations are included in §2.4.

Following Rosado, Sesana & Gair (2015) and Taylor et al. (2016b) we construct simulated PTA using published specifications of the constituent pulsars. We then calculate the resulting DP (Eq. 2.30) for

\* $\Lambda_{ij}$  is equivalent to the multiplicative factors used in, for example, Sesana, Vecchio & Colacino (2008, Fig. 6) and effectively the same as in McWilliams, Ostriker & Pretorius (2014).

<sup>†</sup>i.e. measure of signal strength in PTA (mock) data.

our model GWB against each PTA, focusing on the varying time to detection. We consider models for all PTA:

- European\* (EPTA, Desvignes et al. 2016; Caballero et al. 2016)
- NANOGrav† (The NANOGrav Collaboration et al. 2015),
- Parkes‡ (PPTA, Shannon et al. 2015; Reardon et al. 2016),
- International§ (IPTA, Verbiest et al. 2016; Lentati et al. 2016).

For each array we include an ‘expanded’ model (denoted by ‘+’) including the addition of a pulsar every  $X$  years, where for the individual PTA,  $X = 1/4$ , and for the IPTA,  $X = 1/6$  (Taylor et al. 2016b). All expanded pulsars are given a TOA accuracy of 250 ns, and a random sky location. The public PTA specifications include observation times for each TOA of all pulsars in the array. We take the first TOA as the start time of observations for the corresponding pulsar, and use the overall number of calendar days with TOA measurements¶ to determine the characteristic observing cadence. The cadence from the PTA data files is assumed to continue for the pulsars added in expansion.

Detection statistics depend on pairs of pulsars. For each pair, we set the observational duration as the stretch of time over which both pulsars were being observed:  $T_{ij} = T_i \cap T_j$ ; and take the characteristic cadence as the maximum from that of each pulsar:  $\Delta t_{ij} = \max(\Delta t_i, \Delta t_j)$ . The sensitive frequencies for each pair is then determined by Nyquist sampling with  $\Delta f_{ij} = 1/T_{ij}$ , such that each frequency  $f_{ijk} = 1/T_{ij} + k/\Delta t_{ij}$ , and  $f \in [1/T, 1/\Delta t]$ . In calibrating the  $\lambda_{\text{noise}}$  parameter, we take the end time of observations as the last TOA recorded in the public data files, while for calculations of time to detection, we start with an end point of 2017/01/01.

Pulsars are characterized by a (white-noise) standard deviation  $\sigma_i$  in their TOA. For a time interval  $\Delta t$ , the white-noise power spectrum  $P_{w,i}$  is given by,

$$P_{w,i} = 2\sigma_i^2 \Delta t. \quad (2.17)$$

Some pulsars in the public PTA data provide specifications for a red-noise term|| which we also include. We assume that when red-noise specifications are not provided that they are negligible.

To calibrate our calculations to the more comprehensive analyses employed by the PTA groups themselves, we rescale the white noise ( $\sigma_i$ ) of each pulsar by a factor  $\lambda_{\text{noise}}$  (the procedure described in Taylor et al. 2016b). To determine  $\lambda_{\text{noise}}$ , we calculate upper-limits on the GWB amplitude  $A_{\text{yr}^{-1},\text{ul}}$  (Taylor et al. 2016b, Eq. 4), with a ‘true’ (i.e. injected) GWB amplitude of  $A_{\text{yr}^{-1}} = 0.6 \times 10^{-15}$ , and iteratively adjust  $\lambda_{\text{noise}}$  until the calculated  $A_{\text{yr}^{-1},\text{ul}}$  matches the published values. The total noise used in our calculations is then\*\*,

$$P_i = \lambda_{\text{noise}}^2 P_{w,i} + P_{r,i}. \quad (2.18)$$

\*[http://www.epta.eu.org/aom/EPTA\\_v2.2\\_git.tar](http://www.epta.eu.org/aom/EPTA_v2.2_git.tar)

†<https://data.nanograv.org/>

‡<http://doi.org/10.4225/08/561EFD72D0409>

§[http://www.ipta4gw.org/?page\\_id=519](http://www.ipta4gw.org/?page_id=519)

¶Grouping TOA by observation day deals with near-simultaneous observations at different frequencies which are not representative of the true observing cadence.

||Models for each PTA are: European–Caballero et al. (2016, Eq.3); NANOGrav–The NANOGrav Collaboration et al. (2015, Eq.4); Parkes–Reardon et al. (2016, Eq.4); International–Lentati et al. (2016, Eq.10). Note that for the red-noise amplitudes included in the IPTA public data release, the frequency  $f$  must be given in  $\text{yr}^{-1}$ , and the duration  $T$  in yr (for Eq.10 of Lentati et al. 2016).

\*\*In the presence of red-noise the power spectrum is frequency dependent, i.e.  $P_{ik} = P_i(f_k)$ , but we suppress the additional subscript for convenience.



| Name              | N  | Red | Medians (Pulsars / Pairs)     |              |               | $\lambda_{\text{noise}}$ |
|-------------------|----|-----|-------------------------------|--------------|---------------|--------------------------|
|                   |    |     | $\sigma$<br>[ $\mu\text{s}$ ] | Dur.<br>[yr] | Cad.<br>[day] |                          |
| European          | 42 | 8   | 6.5 / 6.9                     | 9.7 / 8.2    | 14 / 20       | 2.26                     |
| NANOGrav          | 37 | 10  | 0.31 / 0.26                   | 5.6 / 2.3    | 14 / 14       | 3.72                     |
| Parkes            | 20 | 15  | 1.8 / 1.8                     | 15.4 / 9.1   | 21 / 23       | 0.1                      |
| IPTA <sub>0</sub> | 49 | 16  | 3.5 / 3.4                     | 10.8 / 5.8   | 15 / 23       | 5.46                     |
| IPTA'             | 49 | 27  | 1.2 / 1.6                     | 12.8 / 8.2   | 14 / 17       | 1.0                      |

**Table 2.1:** Summary of parameters for the individual and International PTA used in our calculations. The first and second columns give the number of pulsars (N) and the number which include a red-noise model (Red) in the official specifications. The following three columns—the noise ( $\sigma$ ), and observational duration & cadence—are each given as median value for ‘pulsars/pulsar-pairs’. Durations are those up to the end-time of each public data set (ranging from 2011 for Parkes, to 2015 for the EPTA). The observing cadence is calculated based on the total number of days with TOA entries between the first and last recorded observations. The IPTA<sub>0</sub> is based on the official IPTA data release while the IPTA' is a manual combination of specifications from each individual PTA data release, without calibrating to any published upper-limit (i.e.  $\lambda_{\text{noise}} \equiv 1.0$ ), but using the calibration from the individual PTA.

Detailed specifications of each PTA configuration are included online as JSON files. The basic parameters of each individual array and the IPTA are summarized in Table 2.1. The values of  $\lambda_{\text{noise}}$  say something about how consistent the overall noise-parameters are with the upper-limits calculated *in our framework*. Values of  $\lambda_{\text{noise}} > 1$  suggest that additional noise is required.

The International PTA (IPTA<sub>0</sub>) requires the largest  $\lambda_{\text{noise}}$ , suggesting that either the noise is under estimated or that the calculated upper-limit is sub-optimal—possibly due to systematics in combining data from not only numerous telescopes, but numerous groups and/or methodologies. To address this issue we also present our results analyzed against an alternative ‘IPTA'’. The IPTA' is constructed by manually combining the TOA measurements from the individual arrays but without re-calibrating. The total number of pulsars across all PTA is 68, but we only include those from the official IPTA specification (49) as the additional 19 pulsars produce a  $\lesssim 1\%$  improvement in the resulting statistics.

To construct the IPTA', white noise parameters are added in quadrature using the  $\lambda_{\text{noise}}$  from each individual PTA. When a pulsar has multiple red-noise models from different groups, we use the model with the lowest noise power at  $f = 0.1 \text{ yr}^{-1}$ . The red-noise characteristics of the pulsars have a substantial effect on the resulting detection statistics. Because the IPTA' model incorporates the red-noise from all PTA, it ends up being significantly disadvantaged compared to the EPTA (for example) for which only 8 pulsars have red-noise models. To level the playing field, we copy the IPTA' red-noise models to pulsars of each individual PTA which do not otherwise include one. Thus the actual number of pulsars using red-noise models, differs from the values shown in Table 2.1, specifically: our EPTA, NANOGrav, and Parkes models end up with 21, 16, and 18 pulsars with red-noise, and both IPTA models have 27.

Our models for the expanded versions of the individual PTA and IPTA<sub>0</sub> begin adding pulsars after the end of their public data sets, which end in 2015/01 (EPTA), 2013/11 (NANOGrav), 2011/02 (PPTA), & 2014/11 (IPTA<sub>0</sub>). The parameters of the expansion pulsars are low white-noise and zero red-noise, which may give an unfair advantage, for example, to the PPTA vs. the EPTA by adding roughly 16 low-noise pulsars by 2015/01. To try to take this into account for the IPTA'+ model, we add 2 pulsars per year after 2011/02 (when the PPTA data set ends), 4 per year after 2013/11, and finally 6 per year after

2015/01—the same time at which IPTA<sub>0</sub> begins expansion with 6 pulsars per year.

To test our PTA configurations and detection statistics, we reproduce the results of Taylor et al. (2016b, Fig. 2) for purely power-law GWB spectra. Our detection probabilities for each PTA are shown in Fig. 2.13. The left column shows expanded (+) arrays, with new pulsars added each year, while the right column shows the current array configurations. Our Detection Probabilities (DP) are close to those of Taylor et al. (2016b), but not identical, likely due to differences between our noise scaling-factors  $\lambda_{\text{noise}}$ \*. Each panel has vertical lines denoting the year in which each PTA reaches 50% (short-dashed) and 95% DP (long-dashed). For these power-law GWB models at our fiducial  $A_{\text{yr}-1} = 0.6 \times 10^{-15}$  (Paper-1), the individual PTA reach 95% DP in roughly 2031, 2027 & 2027 for the EPTA+, NANOGrav+ & PPTA+ respectively. The IPTA<sub>0</sub>+ reaches the same DP at 2029, while IPTA'+ cuts that down to 2025. IPTA<sub>0</sub>+ performing worse than NANOGrav and Parkes further motivates the usage of the IPTA'+ model. The difference between static and expanding arrays is quite significant. For example, by 2037, none of the static PTA models reach 95% DP. This highlights the importance of continuing to observe known pulsars, while also surveying the sky looking for new ones. Survey programs have been carried out by several large radio telescopes around the world, including GBT and Arecibo in the US, and their continued efforts and funding is of critical importance for PTA science.

## 2.3 RESULTS

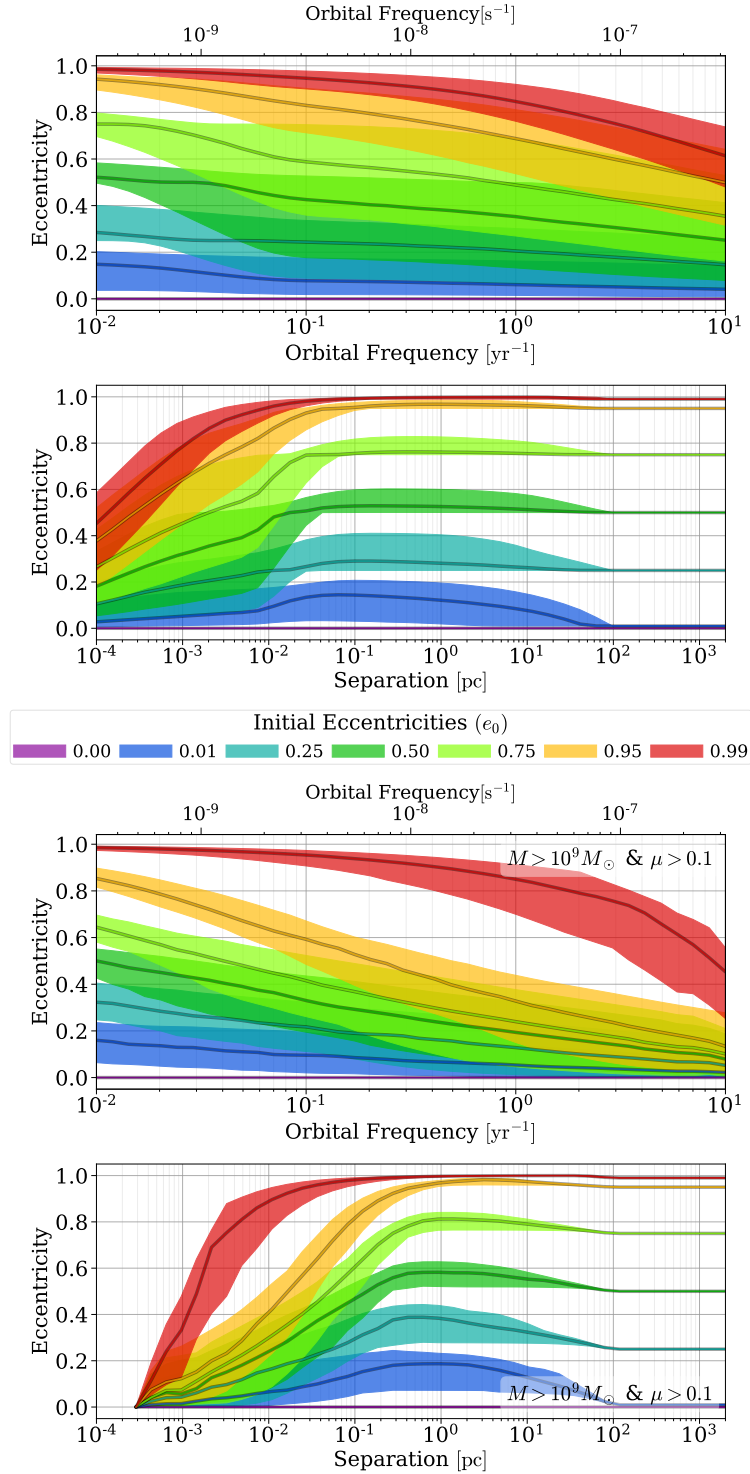
### 2.3.1 ECCENTRIC EVOLUTION

For a given simulation, all binaries are initialized with the same eccentricity and are let to evolve with dynamical friction (DF), loss-cone (LC) stellar scattering, drag from a circumbinary Viscous-Disk (VD), and Gravitational Wave (GW) hardening. LC increases initially-nonzero eccentricities, and GW emission decreases them. In our models we assume DF & VD do not affect the eccentricity distribution. The upper panels of Fig. 2.1 show the resulting binary eccentricity evolution versus orbital frequency (a) and binary separation (b) for our entire population of MBHB. Solid lines show median values at each separation, and colored bands show the surrounding 68%. Binaries initialized to  $e_0 = 0.0$  stay at zero, but the population initialized to only  $e_0 = 0.01$  have a median roughly ten times larger at  $r \sim 1$  pc, and still  $e \gtrsim 0.05$  at orbital frequencies of  $1 \text{ yr}^{-1}$ .

Binaries which are both *heavy* ( $M > 10^9 M_{\odot}$ ) and *major* ( $\mu > 0.1$ ) tend to dominate the GWB signal (Kelley, Blecha & Hernquist 2017). The eccentricity evolution of this subset of binaries is shown separately in panels (c) & (d) of Fig. 2.1. The eccentricities of these systems tend to dampen more quickly at separations below  $\sim 1$  pc. In general, this leads to much lower eccentricities at PTA frequencies for the *heavy* & *major* population. In the  $e_0 = 0.95$  model, for example, the *heavy* & *major* median eccentricity drops below 0.6 by  $f = 0.1 \text{ yr}^{-1}$  (panel c), whereas it takes until  $f \sim 3 \text{ yr}^{-1}$  for the population of all binaries (panel a). In terms of separation, most of the  $e_0 = 0.95$  population has  $e \lesssim 0.5$  by  $r \sim 10^{-2}$  pc, whereas for all binaries the same isn't true until  $r \sim 10^{-4}$  pc. The highest eccentricity model,  $e_0 = 0.99$ , behaves somewhat differently, as it tends to in many respects. For  $e_0 = 0.99$ , binaries pass through most of the PTA band before their eccentricities are substantially damped in both the overall and *heavy* & *major* groups.

Once binaries begin to approach the PTA sensitive band ( $f \gtrsim 10^{-2} \text{ yr}^{-1}$ ) the eccentricity distribu-

\*Taylor et al. (2016b) did not publish their  $\lambda_{\text{noise}}$ , but they were calculated by marginalizing against a probability distribution for  $A_{\text{yr}-1}$  from Sesana (2013b). We simply use a fixed value based on the GWB from our fiducial MBHB merger model,  $A_{\text{yr}-1} = 0.6 \times 10^{-15}$ .



**Figure 2.1:** Evolution of eccentricity versus separation (panels a & c) and orbital frequency (panels b & d) for a variety of initial eccentricities. The upper panels (a & b) show eccentricities for all binaries, while the lower panels (c & d) show only the *heavy* ( $M > 10^9 M_\odot$ ) and *major* ( $\mu > 0.1$ ) subset which tend to dominate the GWB. Each band corresponds to 68% of the population, and the central lines to the medians.

tions are always monotonically decreasing. At the corresponding separations, GW becomes more and more dominant to LC, but at the same time, VD can still be an important hardening mechanism. If circumbinary disks can be effective at increasing eccentricity, or if a resonant third MBH were present, eccentricities could still be excited in the PTA regime. Either of these effects may be important for some MBHB systems, but likely not for the overall population.

In our models, while binary eccentricity is evolved by both LC and GW hardening, the eccentricity distribution itself only affects the rate of *semi-major-axis* hardening (i.e.  $da/dt$ , Eq. 2.3) by the GW mechanism. As eccentricity increases, GW hardening becomes more effective. The hardening time ( $a/[da/dt]$ ) for all binaries is plotted in the upper panels of Fig. 2.2. The frequency panel (a) has been extended to show more of the physical picture, with the PTA-relevant regime shaded in grey. The DF regime goes from large separations down to  $r \sim 50$  pc, and according to our prescription includes no eccentricity dependence.

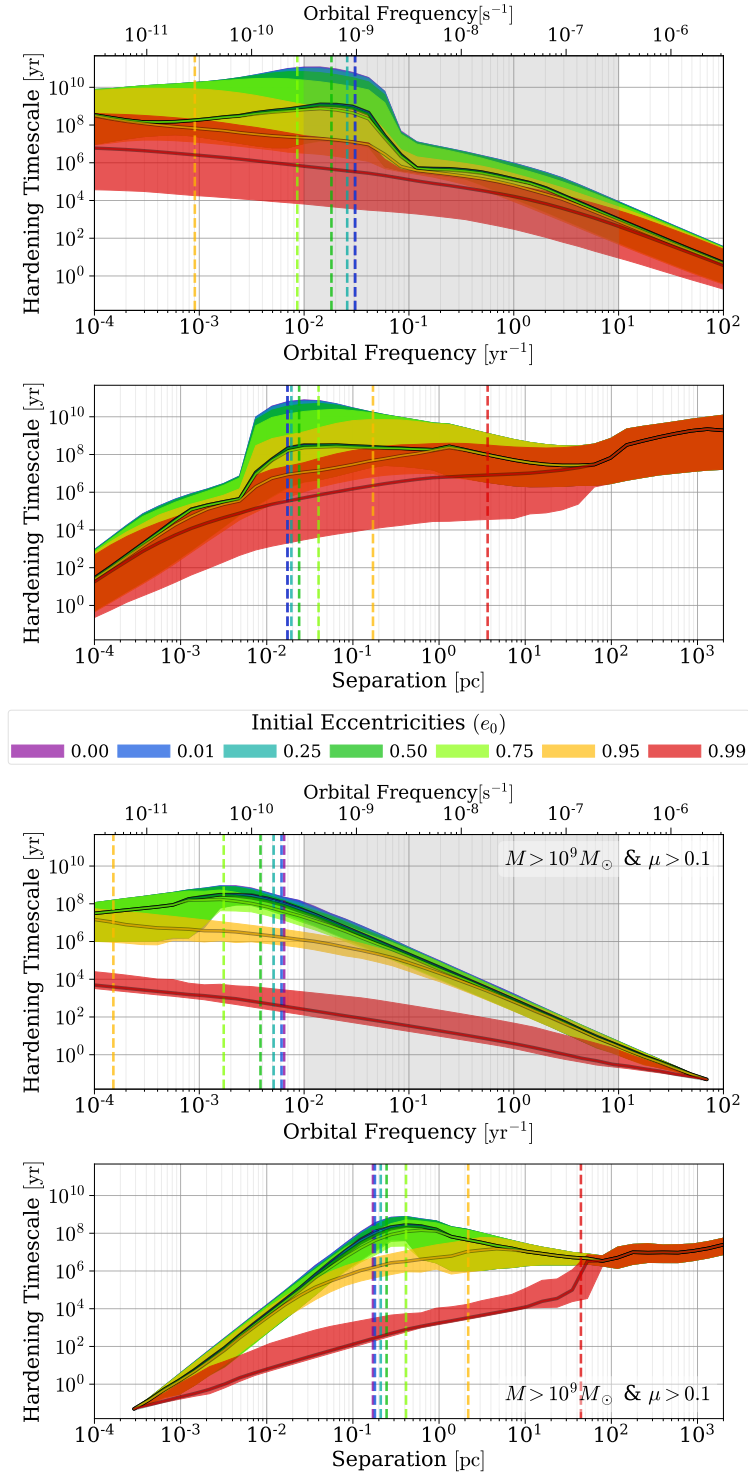
The relatively flat portion of the hardening curve between  $r \sim 50$  pc and  $r \sim 2 \times 10^{-2}$  pc (panel b) is typically the LC-dominated regime. Ignoring the radial dependence of galactic  $\rho/\sigma$  (density/velocity) profiles, the hardening rate should scale like  $\tau_h \propto a^{-1}$  (see Eq. 2.1). The hardening rates in panel b, however, show the combined binary evolutionary tracks over many orders of magnitude in total-mass and mass-ratio, flattening the hardening curves. The scaling is more clear for the low-eccentricity models of the *heavy* & *major* subset (panel d) which show more canonical LC hardening rates between  $r \sim 10$  pc and  $r \sim 1$  pc.

Even at these large separations, the hardening timescale is significantly decreased for the highest initial eccentricities:  $e_0 = 0.95$  and especially  $e_0 = 0.99$ . In these models, GW emission begins to play an important role in hardening much earlier in the systems' evolution. The dashed vertical lines in Fig. 2.2 show the frequency and separation at which GW hardening becomes dominant over DF and LC\* in 50% of systems. For  $e_0 = 0.99$ , GW domination occurs at a few parsecs, largely circumventing the LC regime entirely. As eccentricity decreases, so does the transition separation. For lower initial eccentricities,  $e_0 \lesssim 0.75$ , the DF & LC become sub-dominant at a few times  $10^{-2}$  pc, and the overall hardening rates are hardly distinguishable between different eccentricities.

Panels (c) & (d) of Fig. 2.2 show the hardening time for the population of *heavy* & *major* binaries. While very similar to the overall population, the *heavy* & *major* binaries are all effectively in the GW regime at frequencies  $f \gtrsim 10^{-2} \text{ yr}^{-1}$ , i.e. the entire PTA band. In the  $e_0 = 0.0$  model, binaries in the PTA-band show a nearly perfect power-law hardening rate from GW-evolution, and eccentricities  $e_0 \lesssim 0.75$  are hardly different. In the  $e_0 = 0.95$  model, a heightened hardening rate is clearly apparent up to  $f \sim 10^{-1} \text{ yr}^{-1}$ . The  $e_0 = 0.99$  model evolves five orders of magnitude faster than  $e_0 = 0.0$  at  $f = 10^{-2} \text{ yr}^{-1}$ , and still more than two orders faster even at  $f = 1 \text{ yr}^{-1}$ . The  $e_0 = 0.99$  model hardens drastically faster than the others both because of the strong eccentricity dependence of the GW hardening rate (Eq. 2.3) and because the eccentricity of the  $e_0 = 0.99$  population better retains its high values at smaller separations.

At frequencies below  $10^{-3}$ – $10^{-2} \text{ yr}^{-1}$  and lower eccentricity models ( $e_0 \lesssim 0.75$ ), the *heavy* & *major* population is LC-dominated, causing sharp turnovers in their hardening curves which is echoed in the resulting GWB spectra. In the population of all binaries (panel a) the transition occurs at slightly higher frequencies. The higher eccentricity models, which transition to GW domination at lower frequencies, do not show the same break in their hardening rate evolution. The resulting GWB spectra, however, still do (discussed in §2.3.2).

\*Note that VD can be dominant for an additional decade of frequency or separation, but because the amplitude and power-law index of VD are very similar to that of GW hardening, the transition points plotted are more representative of a change in hardening rate and/or GWB spectral shape.



**Figure 2.2:** Binary hardening timescale ( $a/(da/dt)$ ) against orbital frequency (panels a & c) and binary separation (panels b & d). Note that the frequency axes show an extended range, with the PTA-sensitive band shaded in grey. The upper panels (a & b) show hardening rates for all binaries, while the lower panels (c & d) show only the *heavy & major* subset. The dashed vertical lines indicate the radii at which GW-driven hardening becomes dominant over DF and LC.

In all of our models, GW hardening is dominant to both dynamical friction and stellar scattering for the binaries which will dominate GWB production (*heavy & major* at  $f \gtrsim 10^{-2} \text{ yr}^{-1}$ ). Viscous drag, however, tends to be comparable or higher in much of the binary population. Deviations to the power-law index of hardening rates in the PTA regime are not apparent, however, because VD has a very similar radial dependence to GW emission\* (see §2.2.2). Although the GWB spectrum is only subtly affected by VD, the presence of significant gas accretion in the PTA band is promising for observing electromagnetic counterparts to GW sources, and even multi-messenger detections with ‘deterministic’/‘continuous’ GW-sources—those MBHB resolvable above the stochastic GWB (e.g. Sesana et al. 2012; Tanaka, Menou & Haiman 2012; Burke-Spolaor 2013).

It should be clear from Fig. 2.2 that the hardening timescale is often very long ( $\tau_h \sim \tau_{\text{Hubble}}$ ) and varies significantly between binary systems. A substantial fraction of binaries, especially those with lower total mass and more extreme mass ratios, are unable to coalesce before redshift zero. The presence of significant eccentricity at sub-parsec scales can significantly decrease the hardening timescale, increasing the fraction of coalescing systems overall. The *heavy & major* subset of binaries, however, already tends to coalesce more effectively, and because this portion of the population dominates the GWB, its amplitude due to varying coalescing fractions is only subtly altered.

## 2.3.2 GRAVITATIONAL WAVE BACKGROUNDS

### 2.3.2.1 THE SEMI-ANALYTIC GWB

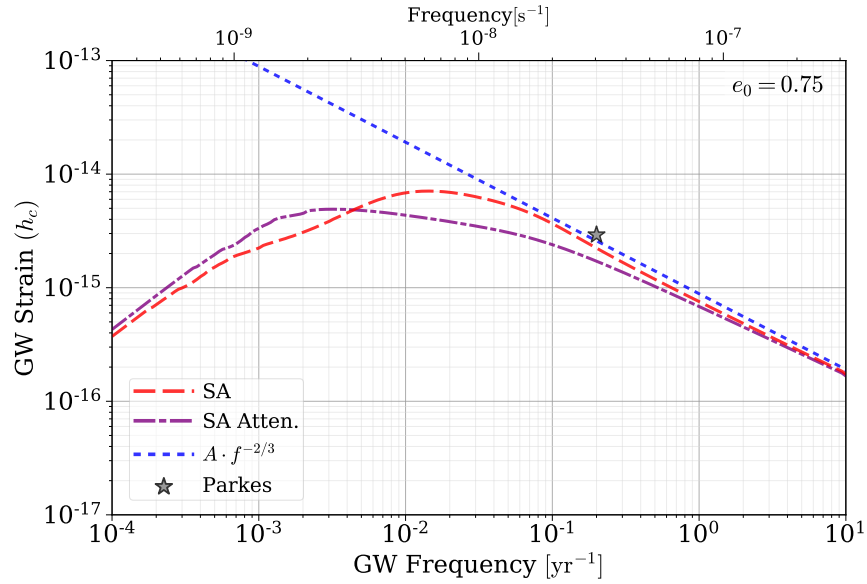
The simplest calculation of the GWB assumes that all binaries coalesce effectively, quickly, and purely due to GW emission. This leads to a purely power-law strain spectrum,  $h_{c, \text{GW}} \propto f^{-2/3}$ . Environmental hardening, however, is required for the vast majority of binaries to be able to coalesce within a Hubble time†. These additional hardening mechanisms also force binaries to pass through each frequency band faster than they otherwise would, decreasing the GW energy emitted in that band, and attenuating the GW background spectrum (e.g. Eq. 2.11, or Ravi et al. 2014).

Non-zero eccentricity *increases* the instantaneous GW luminosity and decreases the hardening time (Eq. 2.3). Also, the GW emission from eccentric binaries is no longer produced monochromatically at twice the orbital frequency, but instead is redistributed, primarily from lower to higher frequencies (Eq. 2.6). The effects of increased attenuation and the frequency redistribution are shown separately in Fig. 2.3 for  $e_0 = 0.75$ . The blue, dotted line shows a power-law spectrum (i.e. GW-only); while the purple, dashed-dotted line includes attenuation from the surrounding medium (predominantly LC-scattering). The red, dashed line includes the effects of chromatic GW emission, in addition to attenuation, which shifts the spectrum to higher frequencies. This is the full, semi-analytic (SA; §2.2.3.1) calculation. The spectral turnover is produced by the environmental attenuation, while the frequency redistribution slightly increases the amplitude in the PTA-regime ( $f \gtrsim 10^{-2} \text{ yr}^{-1}$ ).

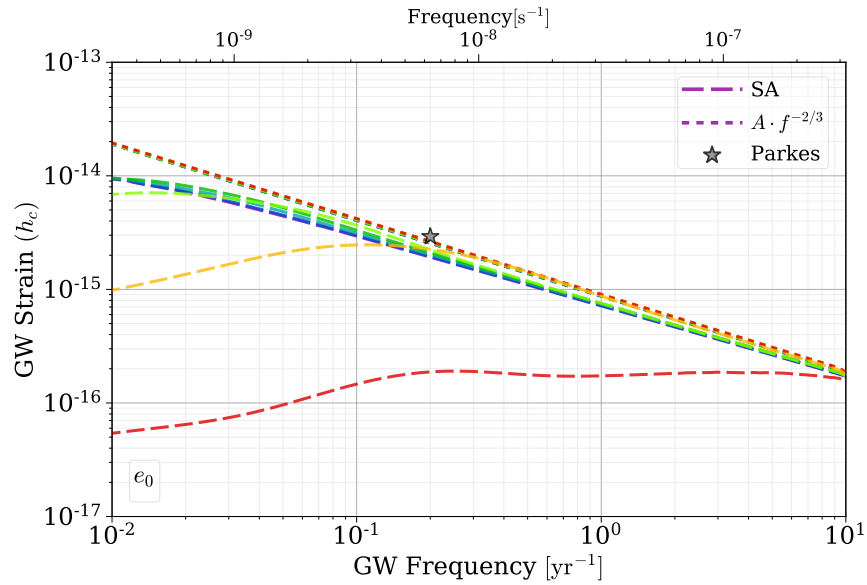
The GWB strain spectrum is shown for a variety of initial eccentricities in Fig. 2.4. Purely power-law spectra calculated assuming GW-only evolution are shown in dotted lines, while the full SA calculation is shown with dashed lines. The power-law spectra show a very slightly increasing GWB amplitude with increasing eccentricity as more binaries are able to coalesce before redshift zero. For initial eccentricities  $e_0 \lesssim 0.5$ , the GWB spectrum above frequencies  $f \gtrsim 10^{-2} \text{ yr}^{-1}$  are nearly identical. For higher initial

\*This is true for the inner, radiation-dominated, region of the disk where GWB-dominating binaries tend to reside.

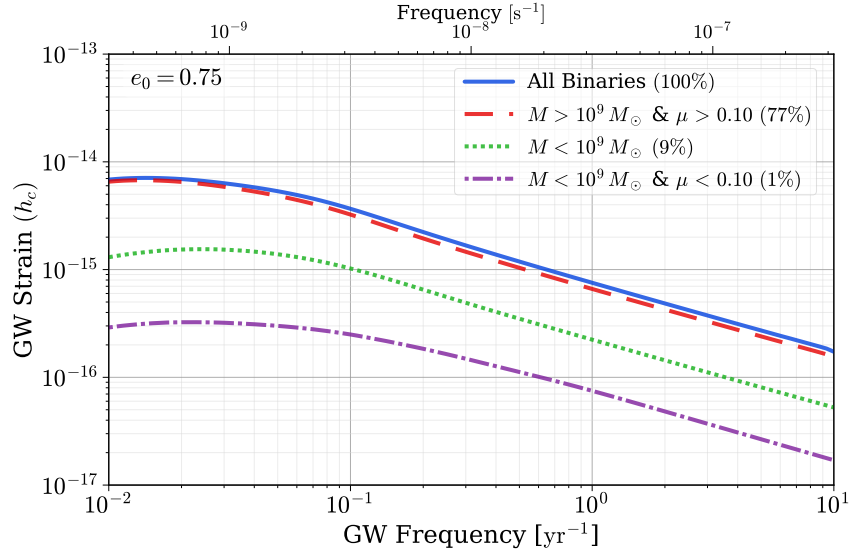
†Even with strong external factors, hardening timescales are still often comparable to a Hubble time, and thus only a fraction of systems coalesce.



**Figure 2.3:** Stochastic Gravitational Wave Background calculated with the Semi-Analytic (SA) method, showing the effects of large binary eccentricities on the GWB spectrum. A purely power-law model is shown in the blue dotted line, which includes the masses and redshifts of merger events but assumes all energy is emitted in GW. The purple dot-dashed line includes the effect of attenuation from environmental hardening effects. The red dashed line shows the full SA calculation including both attenuation and the redistribution of GW energy across multiple frequencies. Note that this model uses a large initial eccentricity of  $e_0 = 0.75$ .



**Figure 2.4:** GWB strain spectrum calculate using the Semi-Analytic (SA) formalism for simulations with a range of initial eccentricities ( $e_0$ ). The dotted lines assume purely GW-driven evolution and no frequency redistribution, and show amplitude increasing slightly with increasing  $e_0$ —because more systems coalesce. The dashed lines show the full SA calculation, including both attenuation and the frequency redistribution. Note that the power-laws (dotted) are not physically meaningful *per se*, and are meant as a reference when interpreting the SA models (dashed). For  $f \gtrsim 2 \times 10^{-1} \text{ yr}^{-1}$ , the GWB amplitude increases with increasing  $e_0$  until  $e_0 = 0.99$ , at which point the GWB amplitude is drastically depressed—almost completely flat—across the PTA band.



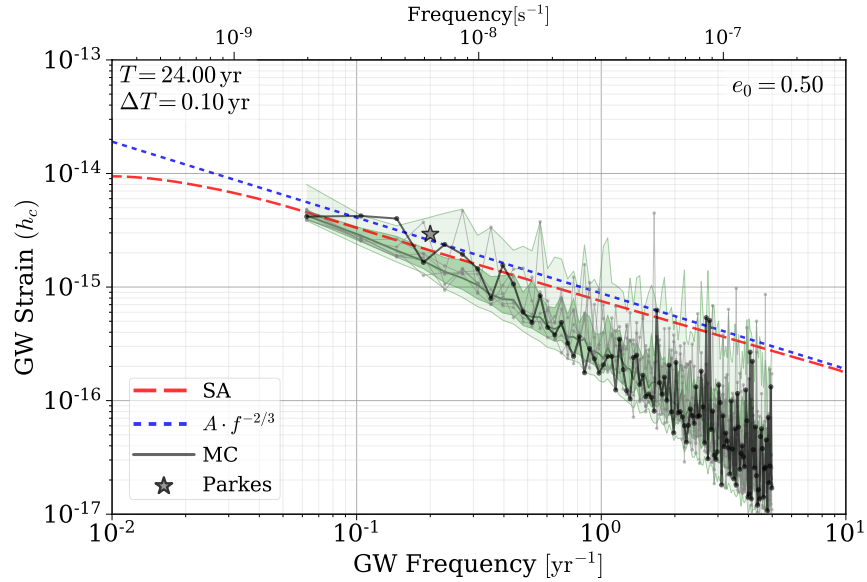
**Figure 2.5:** GWB from different binary subsets selected by total mass  $M$ , and mass ratio  $\mu$ . The blue (solid) line shows the total GWB from all binaries, while the red (dashed) line shows the contribution from *heavy & major* systems (shown also in Figs. 2.1 & 2.2). Light systems (green, dotted) and light & minor binaries (purple, dot-dashed) are also shown. The percentages in the legend show the contribution to the GWB energy at  $1 \text{ yr}^{-1}$  (i.e.  $\propto h_c^2(f = 1 \text{ yr}^{-1})$ ) for each subgroup. Here, the *heavy & major* subgroup, while only 1% of binaries by number, contributes almost 80% of the GWB energy.

values, the effects of non-zero eccentricity become more apparent. Overall, for all but the highest eccentricities ( $e_0 > 0.9$ ), the effect of non-circular orbits is actually to slightly increase the GWB amplitude in the PTA frequency band, because of the redistribution of GW energy to higher harmonics. Note that this is because the spectral turnover is always below the PTA-band in our spectra (also seen in, e.g., [Chen, Sesana & Del Pozzo 2016](#)). If galactic densities are significantly larger than those found in the Illustris simulations, the environment could continue dominating binary evolution to higher frequencies. If that were the case, the spectral turnover could be within the PTA band which would lead to chromatic GW emission *decreasing* the observable signal. In the supplements §2.4, we include additional information on binary host-galaxy densities from Illustris.

In the highest eccentricity case ( $e_0 = 0.99$ ; red lines), binaries are still highly eccentric at separations corresponding to frequencies all the way up to  $f \sim 10 \text{ yr}^{-1}$ , and the GWB amplitude is drastically diminished throughout the PTA band. While such high eccentricities may be unlikely (e.g. [Armitage & Nataraajan 2005](#); [Roedig et al. 2011](#)), having some subset of the population maintain high eccentricities into the PTA band is certainly not impossible (e.g. [Rantala et al. 2016](#)). If Kozai-Lidov-like processes from a hierarchical, third MBH were driving the eccentricity (e.g. [Hoffman & Loeb 2007](#); [Bonetti et al. 2016](#)), or the binary were counter-rotating to the stellar core or circumbinary disk (e.g. [Sesana, Gualandris & Dotti 2011](#); [Amaro-Seoane et al. 2016](#)), eccentricities could grow much faster than in our results. Better understanding what fraction of systems could be susceptible to these processes is an important direction of future study.

The contribution the GWB is broken down by subgroups of total mass and mass ratio in Fig. 2.5. The complete GWB is shown in blue (solid), while the *heavy & major* subset is in red (dashed), the light ( $M < 10^9 M_\odot$ ) in green (dotted), and the light & minor ( $\mu < 0.1$ ) in purple (dash-dotted). The *heavy & major* binaries are only  $\sim 1\%$  of the population, but here, in the  $e_0 = 0.75$  model, contribute  $\sim 80\%$  of the





**Figure 2.6:** GWB calculated using the Monte-Carlo (MC) method, for a simulation with initial eccentricities  $e_0 = 0.5$ . Seven different GWB realizations are shown in gray and black lines, while the median and one- & two- sigma contours are shown in green. Semi-Analytic (SA) calculations are also shown: purely power-law in blue (dotted), and full in red (dashed). The high frequency spectrum of the MC calculation is sharper than that of the SA methods due to quantization of binaries.

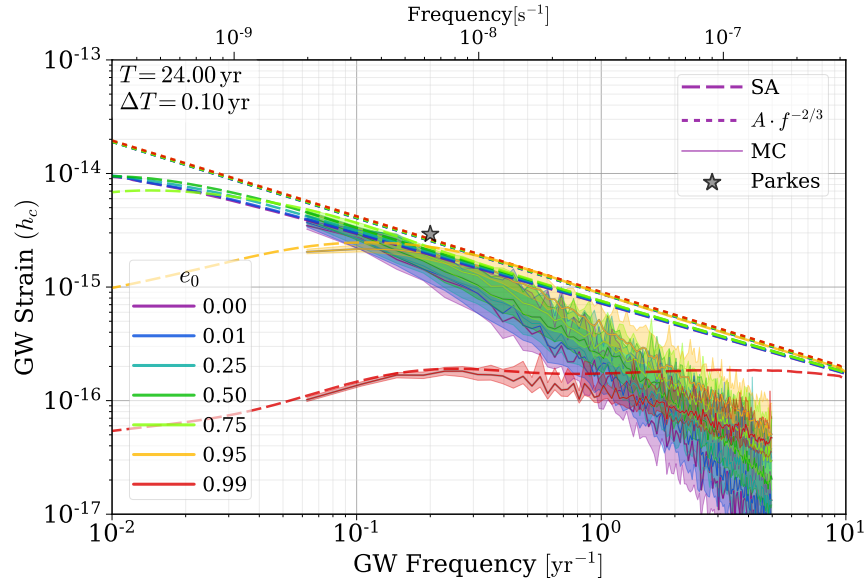
GWB energy density at  $1 \text{ yr}^{-1}$ . In fact, *heavy & major* systems dominate at all frequencies with a very similar spectral shape to the overall GWB, with a slight enhancement of GW strain at lower frequencies. Lower-mass binaries ( $M < 10^9 M_\odot$ ) contribute less than 10% of the energy (despite being just over 90% of the population), and systems which are neither heavy nor massive contribute only 1% ( $\sim 30\%$  of the population). Other initial eccentricity models tend to be even more *heavy & major* dominated (often  $\sim 90\%$ ).

### 2.3.2.2 THE MONTE-CARLO GWB

The Monte-Carlo (MC) approach (§2.2.3.2) to calculating the GWB dispenses with the continuum approximation for the density of GW sources, thereby allowing for the quantization of MBHB at the same time as providing a convenient formalism for constructing an arbitrary number of realizations from the same binary population. The MC GWB is shown in Fig. 2.6, with seven randomly-chosen realizations plotted in gray and black. The median line and one- & two- sigma contours of 200 realizations are shown in green. For reference, the SA spectra are shown for purely power-law calculations (blue, dotted) and the full SA (red, dashed). The frequency bins correspond to Nyquist sampling at a cadence  $\Delta T = 0.1 \text{ yr}$ , and total observational-duration  $T = 24 \text{ yr}$ .

The MC GWB differs from the SA one, both in its jaggedness and also in a steeper spectrum at higher frequencies. The jaggedness is caused by varying numbers of binaries in the observer’s past light cone (Eq. 2.16)—especially massive ones at lower redshifts. The latter effect is due to binary quantization: at high frequencies, where the hardening time is short, there are few MBHB contributing to each frequency bin. The SA calculation, however, implicitly includes the contribution from fractional binaries, artificially inflating the GWB amplitude at high frequencies (Sesana, Vecchio & Colacino 2008).

Figure 2.7 shows the GWB for a variety of initial eccentricities, with median lines and one-sigma contours for the MC calculation, along with the full and power-law SA calculations. Both the SA and MC



**Figure 2.7:** Monte-Carlo (MC) calculated GWB for a variety of different initial eccentricities ( $e_0$ ). Median lines and one-sigma contours are shown for the MC case, and both the full (dashed) and power-law (dotted) SA calculations are also shown. Increasing GWB amplitude with increasing eccentricity is apparent at lower frequencies ( $\sim 10^{-1} \text{ yr}^{-1}$ , until the spectrum turns over. While the spectral turnover is produced by environmental interaction, the frequency at which it occurs is increased with rising eccentricity. At higher frequencies ( $\gtrsim 2 \times 10^{-1} \text{ yr}^{-1}$ ), as eccentricity increases, the MC results come closer and closer to the SA ones due to additional binaries at lower *orbital* frequencies contribute to higher *GW-frequency* bins.

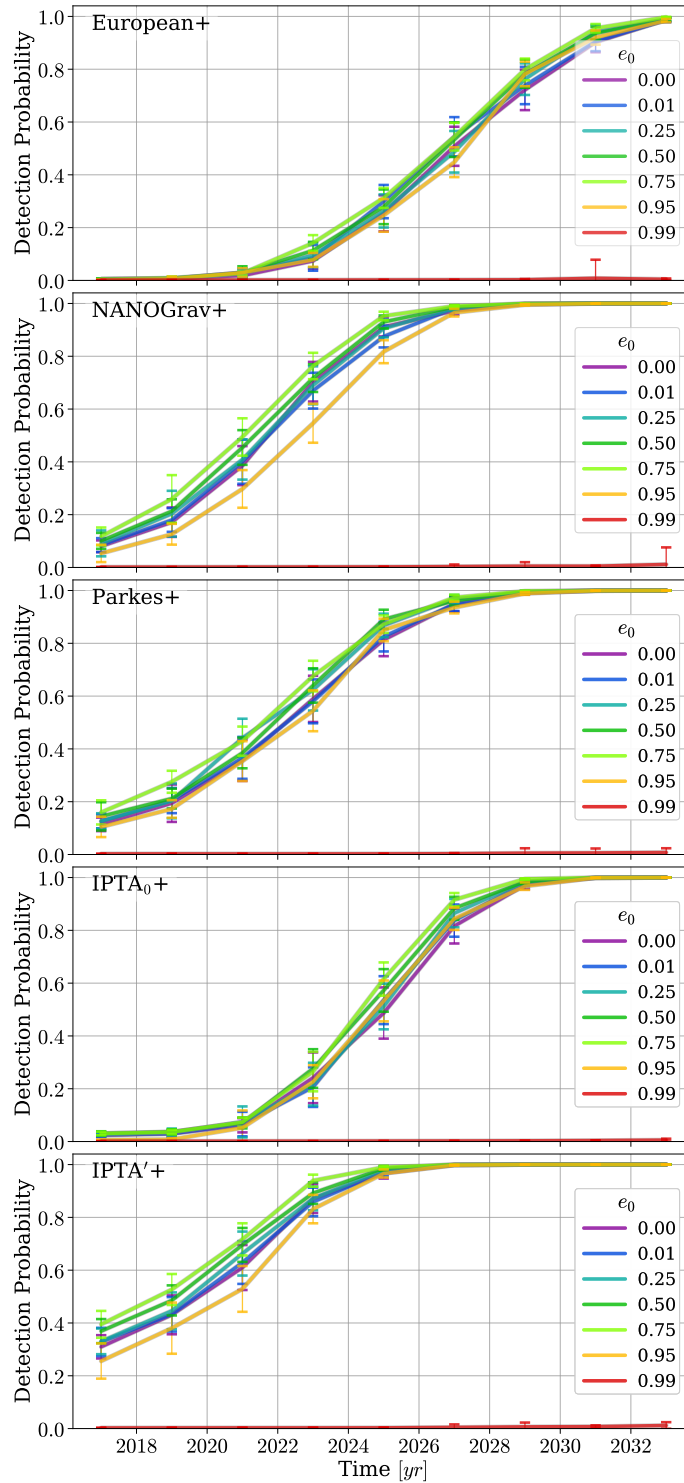
methods show slightly increased GWB amplitudes with increasing eccentricities (except for the highest,  $e_0 = 0.99$  simulation). A more pronounced effect is that the higher the eccentricity, the closer the high-frequency portion of the MC calculation comes to the purely power-law spectra. The ratio of the MC GWB (median lines) to SA GWB is shown explicitly in Fig. 2.14. Higher eccentricities mean that GW energy from binaries at lower *orbital-frequencies* are deposited in higher *GW-frequency* bins. This means that overall, more binaries are contributing to each of the higher frequency bins, reducing the effect of MBHB quantization, and thus bringing the MC results closer in line to the SA models. Finite-number effects at high frequencies are also remediated by increased numbers of coalescing systems, as is the case for very large LC refilling fractions ( $\mathcal{F}_{\text{refill}}$ ), for example, shown in Figs. 2.15 & 2.16.

### 2.3.3 PULSAR TIMING ARRAY DETECTIONS

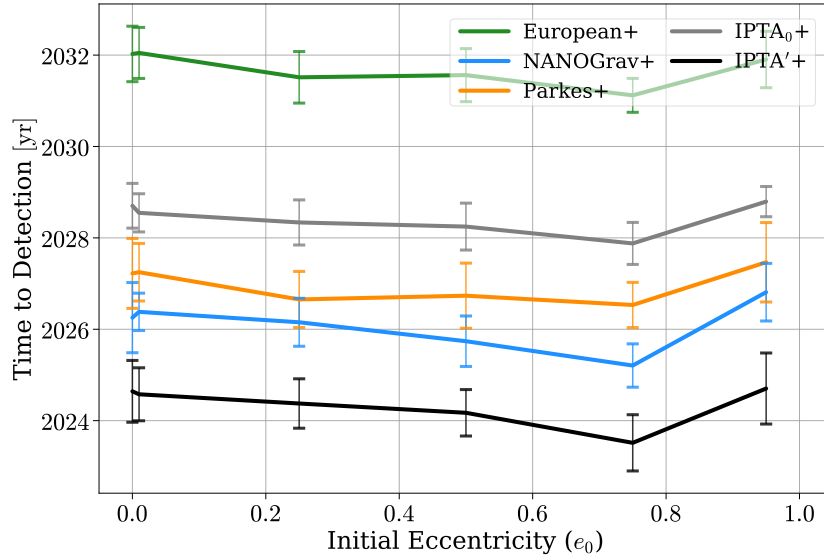
#### 2.3.3.1 ECCENTRIC BINARY EVOLUTION

Detection probability versus time are shown for eccentric-evolution models in Fig. 2.8 using expanded PTA configurations. Averages and standard deviations over 200 MC realizations are plotted. Ignoring the highest eccentricity case ( $e_0 = 0.99$ ), for the moment, the variation between different eccentricities is  $\lesssim 20\%$  in DP, and generally  $\sim 2 \text{ yr}$  at a fixed value. The different DP growth curves are quite similar between PTA, with DP tending to be higher for higher eccentricities between  $e_0 = 0.0$  and  $e_0 = 0.75$ . Not surprisingly, the  $e_0 = 0.99$  model is an outlier in DP as in binary evolution. In this extreme case the GWB is effectively undetectable for all PTA.

The PTA differ in their response to the  $e_0 = 0.95$  simulations depending on their frequency sensitivi-



**Figure 2.8:** Detection Probability (DP) versus time for eccentric evolution models with different initial eccentricities ( $e_0$ , colors) and different PTA (rows). The lines and error bars are averages and standard deviations over 200 MC realizations. Currently, in 2017, we find detection probabilities are below 20% for all *official* PTA, but reach 95% between about 2025 and 2032. Higher eccentricities tend to be slightly more detectable, until  $e_0 > 0.75$  where the spectral turnover takes a toll on the low-frequency GWB amplitude.



**Figure 2.9:** Time to reach a 95% DP versus initial eccentricity ( $e_0$ ) of MBH binaries. Each (expanded, ‘+’) PTA is shown with the average and standard deviation of 200 MC realizations. While the official IPTA<sub>0</sub> specification lags behind Parkes and NANOGrav, the more optimistic IPTA’ model reaches 95% DP almost two years earlier.

ties. Longer observing durations mean PTA are able to detect lower frequencies, and the lowest accessible frequency bins are the most sensitive (see, e.g., Moore, Taylor & Gair 2015). Pulsars (and PTA) with the longest observing durations (i.e. Parkes) are most sensitive to the GWB spectral turnover accentuated in the high eccentricity models. As observing time increases, however, sensitivity increases at all frequencies, and the addition of short duration (and low noise) expansion pulsars boosts high-frequency sensitivity. After a sufficient observing time, the high frequency portion of the GWB spectrum, where  $e_0 = 0.95$  is highest (see Fig. 2.7), gains enough leverage for its DP to overtake that of lower eccentricities.

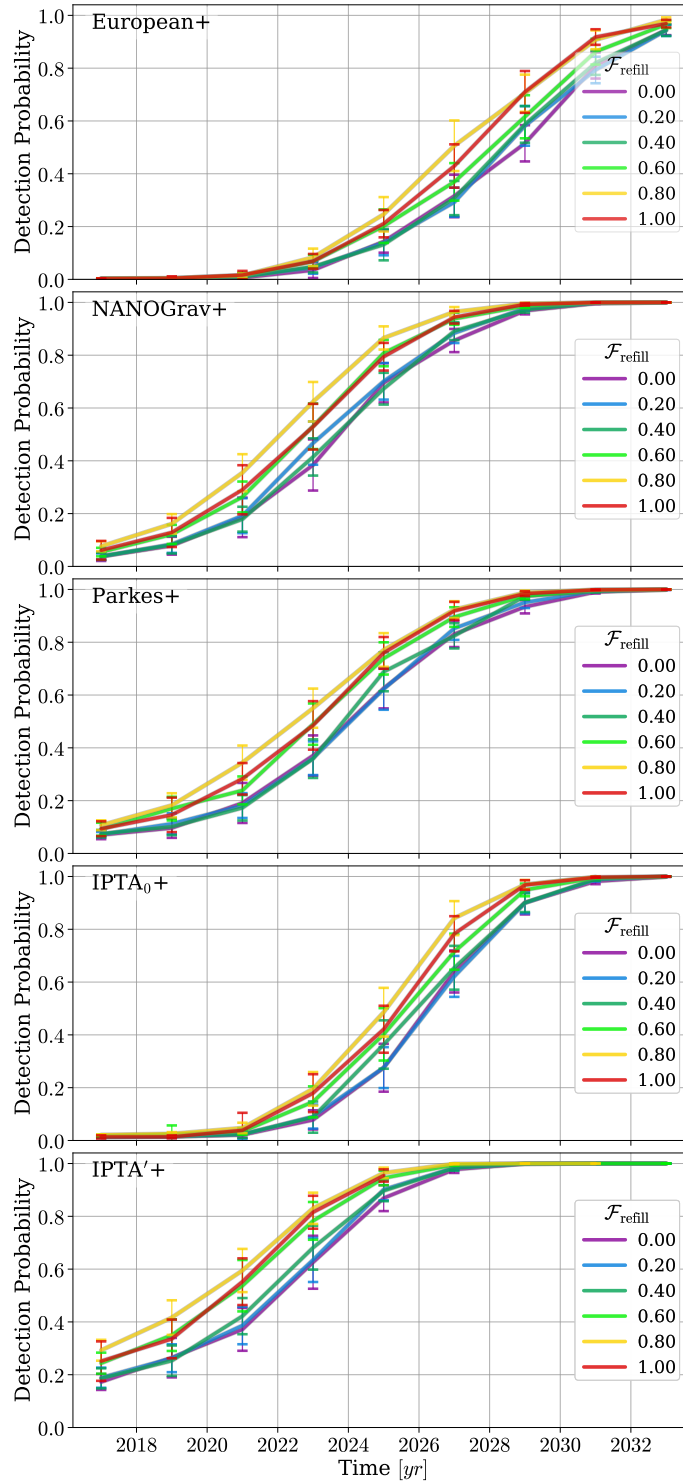
In Fig. 2.8, the IPTA<sub>0</sub>+ doesn’t perform as well as Parkes+ at early times. This is due both to I) the differing noise calibrations—the IPTA<sub>0</sub>’s white-noise is pushed significantly higher than Parkes to match the observed upper-limits, as well as, II) the specifications for each individual PTA not quite matching those included in the IPTA public specifications. Both of these factors motivate our inclusion of the IPTA’ model, which has a higher DP than Parkes, even at early times, as expected. We are optimistic that future IPTA results and data releases will show even greater improvements than suggested by those of the initial IPTA.

Figure 2.9 shows times to reach 95% DP versus initial eccentricity for each PTA. Overall, time to detection tends to improve slightly with increasing eccentricity as it increases the GWB amplitude in the mid-to-upper PTA band. Differences between eccentricities, however, tend to be comparable or smaller than the variance between MC realizations. For very high eccentricities,  $e_0 > 0.75$ , the time to detection again increases as the GWB spectral turnover becomes ‘visible’ to PTA, and the signal is diminished. While the  $e_0 = 0.99$  models never reach 95% DP in our results, additional simulations of a ‘rapid’ IPTA’+ model, where expansion pulsars have a cadence of 2 days (see §2.4, ‘IPTA’<sub>rap</sub>’), are able to reach DP  $\sim 50\%$  by 2032. Even in the highest eccentricity model with  $e_0 = 0.99$ , detection prospects for individual, deterministic sources may not be affected quite as strongly. We are currently exploring single source predictions from our models, to be presented in a future study.

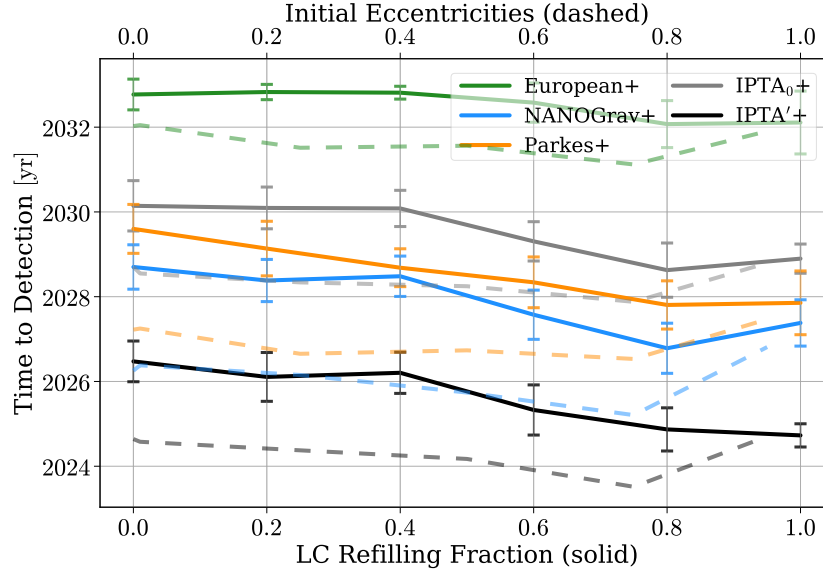
### 2.3.3.2 CIRCULAR BINARY EVOLUTION

Detection probability versus time for *circular* evolution models with a variety of loss-cone refilling parameters ( $\mathcal{F}_{\text{refill}}$ ) are shown in Fig. 2.10. Recall that a low, ‘steady-state LC’ corresponds to  $\mathcal{F}_{\text{refill}} = 0.0$ , and a highly effective, ‘full LC’ to  $\mathcal{F}_{\text{refill}} = 1.0$ . Overall a similar range of durations are required for comparable DP, but the circular models are systematically harder to detect, taking roughly 2 years longer. This is not surprising as the eccentric models assume a full LC, and increasing eccentricity tends to further enhance the GWB amplitude in the PTA band. In general, higher  $\mathcal{F}_{\text{refill}}$  lead to higher DP after a fixed time. The circular, full LC, tends to have a slightly lower DP as the attenuation and spectral turnover from stellar scattering take effect, analogous to the highest eccentricities (see, e.g., Fig. 2.15). The IPTA’+ model shows a slight improvement in time to detection between  $\mathcal{F}_{\text{refill}} = 0.8$  and  $\mathcal{F}_{\text{refill}} = 1.0$ , suggesting that its high-frequency sensitivity is able to win out.

Figure 2.11 summarizes the circular-evolution times to detection for each PTA versus  $\mathcal{F}_{\text{refill}}$  (solid lines). Overplotted are the eccentric-evolution times to detection (dashed lines, upper x-axis) for comparison. Based on these trends, a population of MBHB with very high LC scattering efficiency ( $\mathcal{F}_{\text{refill}} \sim 0.8 - 1.0$ ) and intermediate eccentricities  $0.5 < e_0 < 0.8$  would be the easiest to detect.



**Figure 2.10:** DP versus time for circular evolution models with different LC efficiencies ( $\mathcal{F}_{\text{refill}}$ , colors) and different expanded PTA (rows). These growth curves behave very similarly to those of the eccentric models, but generally take  $\sim 2$  yr longer to reach the same DP. There is also a stronger trend across  $\mathcal{F}_{\text{refill}}$  compared to  $e_0$ , with  $\mathcal{F}_{\text{refill}} = 0.00$  reaching the same DP  $\sim 2$  yr slower than for  $\mathcal{F}_{\text{refill}} \sim 0.8 - 1.0$ .



**Figure 2.11:** Time to reach a 95% DP versus LC refilling fraction ( $\mathcal{F}_{\text{refill}}$ ; solid lines) for circular binary-evolution models. The times to detection for varying eccentricity models (from Fig. 2.9) are overlotted (dashed lines) for comparison. Averages and standard deviations are shown. Varying  $\mathcal{F}_{\text{refill}}$  has a pronounced effect on detection times, with more effective scattering (larger  $\mathcal{F}_{\text{refill}}$ ) models taking less time to observe. Because the eccentric evolution models assume a full loss-cone, the  $\mathcal{F}_{\text{refill}} < 1.0$  models tend to take longer to be detected.

## 2.4 CONCLUSIONS

This paper has focused on plausible detections of a stochastic Gravitational Wave Background (GWB) using Pulsar Timing Arrays (PTA). We have expanded on the Massive Black-Hole (MBH) merger models presented in Kelley, Blecha & Hernquist (2017) based on the Illustris simulations. We have added a model for eccentric binary evolution assuming ‘full’ Loss-Cone (LC) stellar scattering, in addition to our existing prescriptions for dynamical friction, stellar scattering with a variety of LC efficiencies ( $\mathcal{F}_{\text{refill}}$ ), viscous drag from a circumbinary disk, and Gravitational Wave (GW) emission. We have run sets of simulations with a variety of LC efficiencies and initial (at the start of stellar scattering) eccentricities. The MBH binary evolution produced by our models is explored along with Monte-Carlo (MC) realizations of the resulting stochastic GW Background (GWB) spectra. Using parametrized models of currently operational PTA, and their future expansion, we calculate realistic prospects for detections of the GWB.

The presence of non-zero eccentricity causes two distinct effects to MBH Binary (MBHB) evolution and their GW spectra. First, increased eccentricity causes faster GW-hardening and thus more binary coalescence, but there is additional attenuation of the GWB spectrum and a stronger spectral turnover at low frequencies. Second, while circular binaries emit GW at only twice the orbital frequency (the  $n = 2$  harmonic), eccentric binaries also emit at all higher harmonics (and the  $n = 1$ ). The total GW energy released remains the same, but the overall effect is to move GW energy from lower to higher frequencies (Fig. 2.3).

GWB spectra constructed using a Semi-Analytic (SA) calculation (Fig. 2.4) show that the amplitude ( $A_{\nu^{-1}}$ ) of the GWB near the middle of the PTA band ( $f \sim 1 \text{ yr}^{-1}$ ) tends to increase with increasing eccentricity up to  $e_0 = 0.95$ . This is due primarily to the first eccentric effect: with hardening more effective, the number of binaries coalescing by redshift zero increases. At lower frequencies, environmental

effects—specifically stellar scattering—produce a strong turnover in the GWB spectrum. Even moderate eccentricities begin to increase the frequency at which this turnover occurs because of the second eccentric effect. Unless the population of binaries dominating the GWB have very high eccentricities ( $e \gtrsim 0.8$ ), the spectral turnover remains below the PTA sensitive band ( $f \lesssim 0.02 \text{ yr}^{-1}$ ).

The location of the spectral turnover in our models differs from those predicted by [Ravi et al. \(2014\)](#) who see the turnover at frequencies as high as  $f \sim 10^{-1} \text{ yr}^{-1}$ . The location of the turnover depends on the strength of environmental factors, and thus galactic density profiles. If the stellar densities of massive-MBHB host-galaxies are higher than predicted by Illustris, the turnover could occur at PTA-observable frequencies, regardless of (or exaggerated by) eccentricity distribution. If the turnover does exist in the current PTA band, it could hurt detection prospects. At the same time, observations of such a turnover would be uniquely indicative of environmental interactions, while observations of the GWB amplitude overall are highly degenerate between cosmological factors (i.e. the rate of binary formation) and environmental factors (determining the rate of binary coalescence). Currently PTA upper limits of the GWB are still entirely consistent with our results, and thus are unable to constrain or select between them.

GWB spectra constructed using the MC calculation (Fig. 2.7), which should resemble real signals, tend to have much steeper strain spectra than  $-2/3$  at current PTA frequencies ( $f \gtrsim 0.1 \text{ yr}^{-1}$ ). This is because the number of binaries in each frequency bin becomes small, and binary quantization must be taken into account. MC realizations reveal an interesting corollary to the redistribution of GW energy: with non-zero eccentricity, a larger number of binaries at lower *orbital-frequencies* contribute to the GWB signal at higher *observed-frequency* bins. This softens the effect of binary quantization and the GWB spectra tend to come closer and closer to a  $-2/3$  spectral index with increasing eccentricity (Fig. 2.14)—thus producing higher  $A_{\text{yr}^{-1}}$ . For example, the  $e_0 = 0.5$  &  $e_0 = 0.95$  models have  $A_{\text{yr}^{-1}}$ , 2 & 3 times larger than that of the  $e_0 = 0.0$  model.

To calculate realistic detection statistics, we use parametrized version of each operational PTA: the European (EPTA), NANOGrav, Parkes (PPTA), and International (IPTA; a joint effort of the individual three). For the IPTA we consider both the public data specifications, IPTA<sub>0</sub>; in addition to a more optimistic, manual combination of the individual groups, IPTA'. Overall, our models for NANOGrav, Parkes, and IPTA<sub>0</sub> each behave comparably, reaching 95% detection probability (DP) between 2026 and 2030, and the EPTA following 2 – 4 years later. The IPTA' model noticeably outperforms the others, reaching 95% DP between 2024 and 2026. High cadence observations of IPTA' pulsars can further decrease time-to-detection by another  $\sim 2$  years. Moderately high eccentricities ( $0.5 \lesssim e_0 \lesssim 0.8$ ) tend to produce the largest GWB amplitudes in the PTA band. The eccentricity models used here assume a full loss-cone (LC;  $\mathcal{F}_{\text{refill}} \approx 1.0$ ), and thus circular evolution models which decrease the LC refilling efficiency tend to have lower GWB amplitudes, and longer times to detection by up to  $\sim 2$  years. If galactic-nuclear stellar densities are significantly higher than suggested by Illustris, and the LC is also nearly full, then attenuation of the GWB spectrum could increase times to detection.

Increased eccentricity tends to increase the GWB amplitude, and thus detection prospects. In the most extreme  $e_0 = 0.99$  model, however, the signal is so drastically diminished that detections seem unlikely within 20 years in all but the high-cadence IPTA model. While eccentricities as high as 0.99 may not be representative of the overall population of binaries, mechanisms which can preferentially drive more massive systems to higher eccentricities (e.g. counter-rotating stars/gas, [Khan, Just & Merritt 2011](#); [Amaro-Seoane et al. 2016](#), or three-body resonances) should be further studied. Varying the eccentricity distribution of binaries has a strong effect on the strain spectral-index of the GWB at high frequencies ( $f \gtrsim 1 \text{ yr}^{-1}$ ). While PTA are less sensitive at higher frequencies, eventual observations with high signal-to-noise could be used to constrain the underlying binary eccentricity distribution. The true eccentricity



distribution will also affect the prospects for observations of individual, resolvable binaries (‘deterministic’/‘continuous’ sources)—a study of which is currently in progress.

As discussed above, the high frequency portion of the GWB spectrum that PTA will eventually observe is strongly influenced by individual MBHB sources. In effect, the high frequency portion of the spectrum is no longer a ‘background’. It has been shown that this leads to non-Gaussian signal statistics (Ravi et al. 2012) which are at odds with the assumptions of the detection statistics we use. Recently, Cornish & Sampson (2016) have found the standard analyses to be robust against small numbers of GW sources. None the less, if this effect were to systematically decrease GWB detection probabilities, it is likely the effect would be minor because: 1) detection probability is primarily driven at low-frequencies where individual sources are much less important; and 2) the Monte Carlo realization of the spectra we construct should be representative of variations in the GW *background* (neglecting single-sources), and thus our DP and time-to-detection error bars should still be representative.

For a given PTA configuration, the differences in times-to-detection for varying GWB model parameters are at most a few years. This result is promising as it suggests that, despite uncertainties in the underlying physical processes of binary mergers, the expectation of GWB detections in the near future remains robust. At the same time it begs the question, ‘will PTA be able to discern between different models in their observations?’ Based only on the *overall GWB amplitude* (or equivalently the time-to-detection), only a mixed measurement of the overall merger process and the typical MBH binary mass distribution will be constrained. The different hardening models are largely degenerate in the overall GWB amplitude they predict, especially when taking into account uncertainties in cosmological factors—most notably the true, unbiased distribution of MBH masses (e.g. Shen et al. 2008; Shankar et al. 2016)—which is outside of the scope of this study\*.

Once PTA have detected the GWB, and signal-to-noise continues to grow, the *shape* of the GWB will be measured which encodes very detailed information about the merger process and typical MBH environments (e.g. Taylor, Simon & Sampson 2017; Chen et al. 2017). The *strength* of the (low-frequency) spectral turnover is determined by the MBHB coupling to their local stellar environments, and its *location* is additionally effected by the eccentricity distribution of binaries. The (high-frequency) spectral-index, however, measures the number of sources contributing to the GWB, and thus the underlying eccentricity distribution. In the ideal, high signal-to-noise regime, the spectral index will determine typical binary eccentricities which can then be disentangled from the stellar coupling, measured from the spectral turnover. With eccentricity and the loss-cone constrained, the typical masses of merging MBHB can then be inferred from the overall GWB amplitude.

Low frequency sensitivity, established by long observing baselines, tends to drive increases in detection probability. Still, we find that including short cadence observations to maintain or improve high frequency sensitivity can make a noticeable difference in detection prospects, especially for the most extreme hardening and eccentricity models (in which the GWB spectra turn over at low frequencies). Regardless of cadence, we find that the continued addition of pulsars monitored by PTA is *essential* for a detection to be made within the next 20 years. Across a wide range of specific configurations, and even with pessimistic model parameters, if PTA continue to expand as they are, GWB detections are highly likely within about 10 years.

---

\*Examining the effects of varying Illustris MBH evolution and masses is being examined. The MBH population from Illustris—especially at the high-mass end which most strongly effects the GWB—is tightly constrained by the  $M$ - $\sigma$  relation and AGN luminosity function which are both accurately reproduced (Sijacki et al. 2015).

## ACKNOWLEDGMENTS

We are grateful to Pablo Rosado who was extremely helpful in clarifying details of the PTA detection statistics, and to the referee for thorough and wholly constructive feedback which significantly improved this paper.

This research made use of `Astropy`, a community-developed core Python package for Astronomy ([Astropy Collaboration et al. 2013](#)), in addition to `SciPy` ([Jones et al. 2001](#)), `ipython` ([Prez & Granger 2007](#)), `NumPy` ([van der Walt, Colbert & Varoquaux 2011](#)). All figures were generated using `matplotlib` ([Hunter 2007](#)).

## SUPPLEMENTAL MATERIAL

## ADDITIONAL EQUATIONS

The GW frequency distribution function can be expressed as (Peters & Mathews 1963, Eq. 20),

$$g(n, e) \equiv \frac{n^4}{32} \left( [G_1]^2 + [1 - e^2] [G_2]^2 + \frac{4}{3n^2} [J_n(ne)]^2 \right),$$

$$G_1(n, e) \equiv J_{n-2}(ne) - 2eJ_{n-1}(ne) + \frac{2}{n}J_n(ne) + 2eJ_{n+1}(ne) - J_{n+2}(ne), \quad (2.19)$$

$$G_2(n, e) \equiv J_{n-2}(ne) - 2eJ_n(ne) + J_{n+2}(ne).$$

Here  $J_n(x)$  is the  $n$ 'th Bessel Function of the first kind. The sum of all harmonics,  $\sum_{n=1}^{\infty} g(n, e) = F(e)$ , where  $F(e)$  is defined in Eq. 2.4.

The observed, characteristic strain from a set of individual sources is (e.g. Rosado, Sesana & Gair 2015, Eq. 8),

$$h_c^2 = \sum_i h_{s,i}^2 \frac{f_i}{\Delta f} \approx \sum_i h_{s,i}^2 f_i T, \quad (2.20)$$

where the second equality assumes that frequency bins are determined by the resolution corresponding to a total observational duration  $T$ .

The cosmological evolution function is (Hogg 1999, Eq. 14),

$$E(z) \equiv \sqrt{\Omega_M (1+z)^3 + \Omega_k (1+z)^2 + \Omega_\Lambda}, \quad (2.21)$$

for  $z$  the redshift, and  $\Omega_M$ ,  $\Omega_k$  &  $\Omega_\Lambda$  the density parameters for matter, curvature and dark-energy.

## DETECTION FORMALISM

In what follows, the GWB signal is characterized by a Spectral Energy Density (SED),

$$S_h = \frac{h_c^2}{12\pi^2 f^3}, \quad (2.22)$$

and the prediction/model SED is denoted as  $S_{h0}$ . In all of our calculations, we use a purely power-law GWB spectrum for  $S_{h0}$  with an amplitude of  $A_{yr^{-1}} = 10^{-16}$ . Each pulsar  $i$  is characterized by a noise SED  $P_i$  (Eq. 2.18).

PTA detection statistics typically rely on cross-correlations between signals using an ‘overlap reduction function’ (the Hellings & Downs 1983, curve),

$$\Gamma_{ij} = \frac{3}{2}\gamma_{ij} \ln(\gamma_{ij}) - \frac{1}{4}\gamma_{ij} + \frac{1}{2} + \frac{1}{2}\delta_{ij} \quad (2.23)$$

where,

$$\gamma_{ij} = \frac{1}{2} [1 - \cos(\theta_{ij})], \quad (2.24)$$

for an angle between pulsars  $i$  and  $j$ ,  $\theta_{ij}$ .

We employ the ‘B-Statistic’ from Rosado, Sesana & Gair (2015), constructed by maximizing the statistic’s SNR—defined as the expectation value of the statistic in the presence of a signal,

$$\mu_{B1} = 2 \sum_k \sum_{ij} \frac{\Gamma_{ij}^2 S_h S_{h0}}{(P_i + S_{h0})(P_j + S_{h0}) + \Gamma_{ij}^2 S_{h0}^2}, \quad (2.25)$$

divided by the variance of the statistic also in the presence of a signal,

$$\sigma_{B1}^2 = 2 \sum_k \sum_{ij} \frac{\Gamma_{ij}^2 S_{h0}^2 [(P_i + S_h)(P_j + S_h) + \Gamma_{ij}^2 S_h^2]}{[(P_i + S_{h0})(P_j + S_{h0}) + \Gamma_{ij}^2 S_{h0}^2]^2}, \quad (2.26)$$

i.e.  $S/N_B \equiv \mu_{B1}/\sigma_{B1}$ , as apposed to the variance in the *absence* of a signal,

$$\sigma_{B0}^2 = 2 \sum_k \sum_{ij} \frac{\Gamma_{ij}^2 S_{h0}^2 P_i P_j}{[(P_i + S_{h0})(P_j + S_{h0}) + \Gamma_{ij}^2 S_{h0}^2]^2}. \quad (2.27)$$

The SNR can then be expressed as,

$$S/N^2 = S/N_B^2 = 2 \sum_k \sum_{ij} \frac{\Gamma_{ij}^2 S_h^2}{P_i P_j + S_h (P_i + P_j) + S_h^2 (1 + \Gamma_{ij}^2)}, \quad (2.28)$$

which is only meaningful compared to the threshold-SNR for a particular false-alarm probability ( $\alpha_0$ ) and DP-threshold ( $\gamma_0$ ),

$$S/N_B^T = \sqrt{2} \left[ \frac{\sigma_0}{\sigma_1} \operatorname{erfc}^{-1}(2\alpha_0) - \operatorname{erfc}^{-1}(2\gamma_0) \right]. \quad (2.29)$$

The SNR can be circumvented altogether by considering the measured DP,

$$\gamma_B = \frac{1}{2} \operatorname{erfc} \left( \frac{\sqrt{2} \sigma_0 \operatorname{erfc}^{-1}(2\alpha_0) - \mu_1}{\sqrt{2} \sigma_1} \right), \quad (2.30)$$

which is the primary metric we use throughout our analysis of PTA detections.

## HOST-GALAXY DENSITIES IN ILLUSTRIS

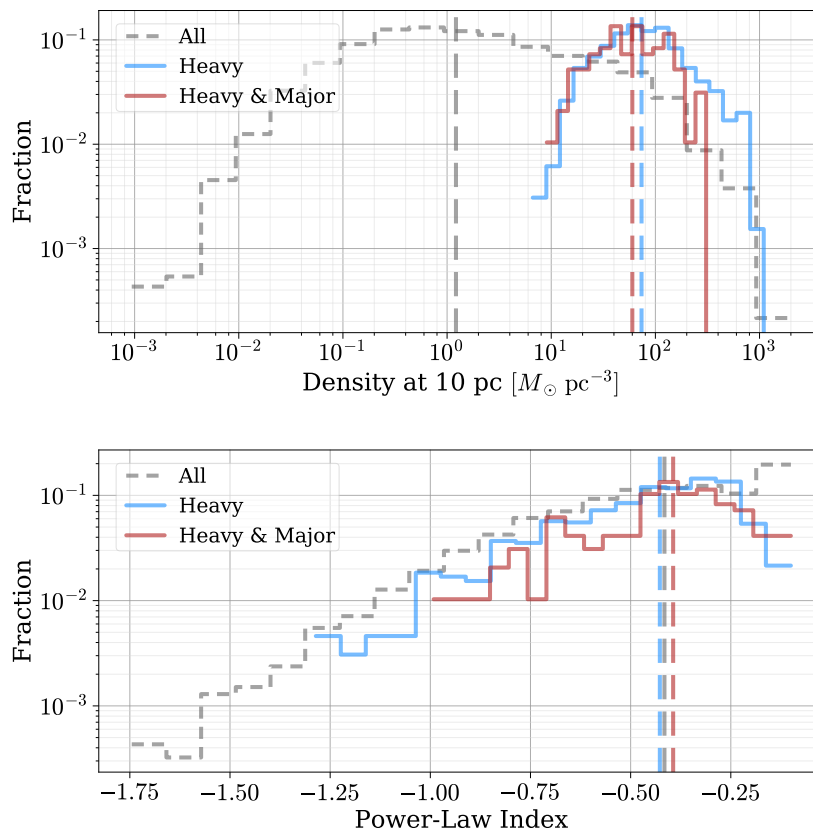
The point at which two MBH come within a smoothing length of one another is identified in Illustris, and density profiles are calculated for the host galaxy at that time. The profiles are used to calculate the environmental hardening rates which then determine the GWB spectra. In particular, the stellar densities strongly affect the location of the spectral turnover through stellar scattering. Because the location of the spectral turnover is especially important for future detections of the GWB, we provide some additional details on the stellar environments here.

To calculate density profiles, we average the density of each particle type (star, dark matter, and gas) in radial bins. Because Illustris is only able to resolve down to 10s–100s of parsec scales, we extrapolate to smaller radii with power-law fits to the eight inner-most bins\*. Fig. 2.12 shows the distribution of stellar densities at 10 pc (interpolated or extrapolated as needed) in the upper-panel†, and power-law indices in the lower-panel. The overall population of binaries are shown in grey (dashed), in addition to the *heavy* ( $M > 10^9 M_\odot$ ) subset in blue, and *heavy & major* ( $\mu > 0.1$ ) mass-ratio subset in red. There is a roughly 100 times increase in typical stellar densities between *heavy* systems and overall host-galaxies, but no noticeable change when further selecting by mass-ratio. While the *heavy* subset constitutes less than 10% of systems, they contribute  $\sim 90\%$  of the GWB amplitude (see, Paper-1).

The median power-law index for the inner stellar density profiles is  $\sim -0.4$ . For comparison, at small radii an Hernquist (1990) profile corresponds to  $-1$ , and  $-1.5$  produces a surface-density distribution that resembles a de Vaucouleurs (1948) profile (Dehnen 1993). At the same time, many massive galaxies (comparable to our host galaxies) have flattened ‘cores’ in their stellar density profiles (e.g. Faber et al. 1997; Lauer et al. 2007) and it has long been proposed that these cores could be explained by dynamical scouring from MBH binaries (e.g. Quinlan & Hernquist 1997; Volonteri, Madau & Haardt 2003). Both computationally and observationally, inner density profiles in the ‘hard’ binary regime (typically  $r \lesssim 10$  pc) are very difficult to resolve. It is thus unclear how accurate these profiles are. While they may be realistic models, some of the flattening in the inner regions may be due in part to numerical effects (e.g. gravitational softening in the force calculations) or the known, over-inflated radii of some galaxies in Illustris

\*Restricted to those which contain at least four particles each.

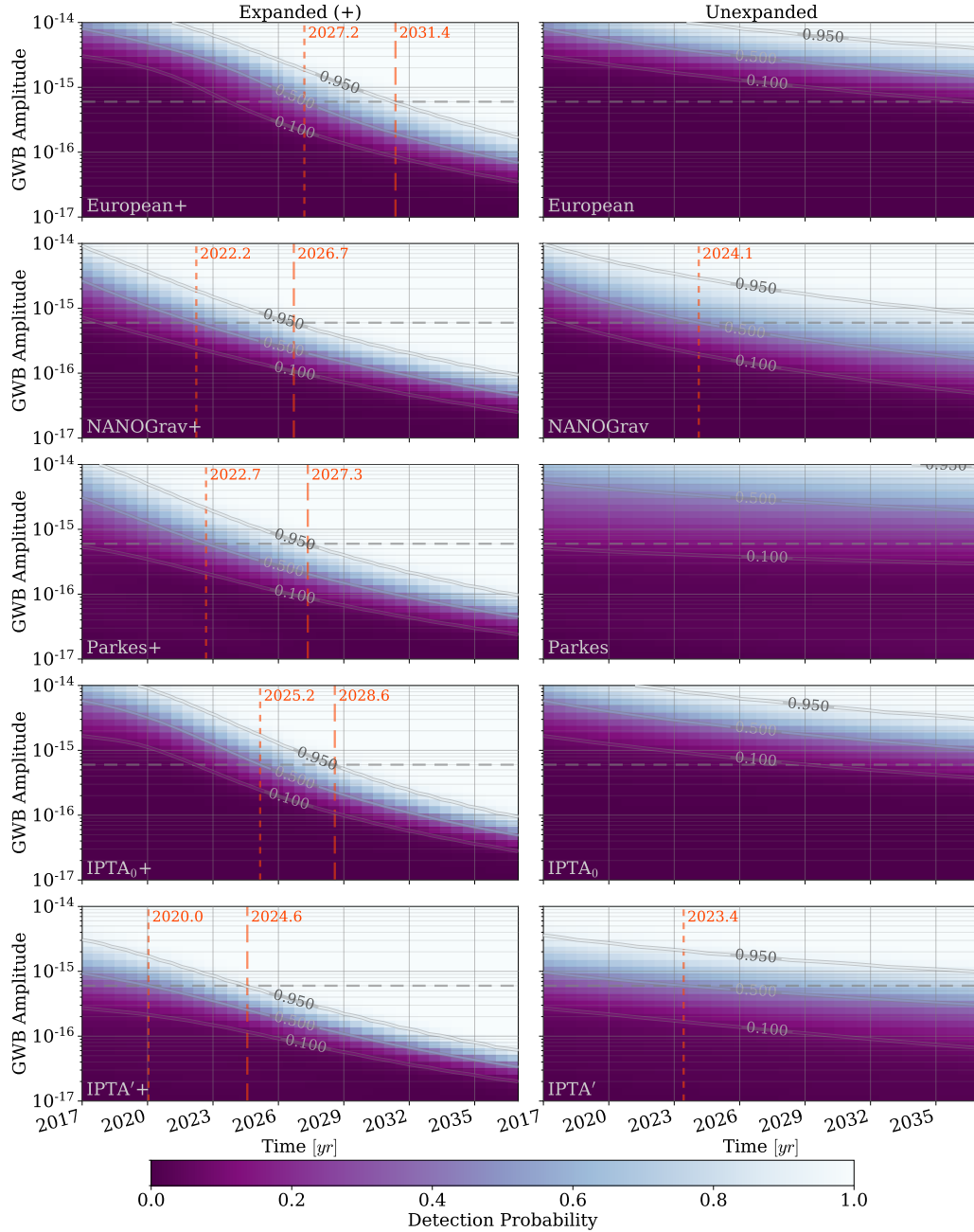
†We choose 10 pc as it is near typical spheres of influence ( $\mathcal{R}_{\text{infl}}$ ) & hardening radii ( $R_h$ ) for our systems, observational resolution-limits for nearby galaxies, and usually just beneath Illustris resolution-limits.



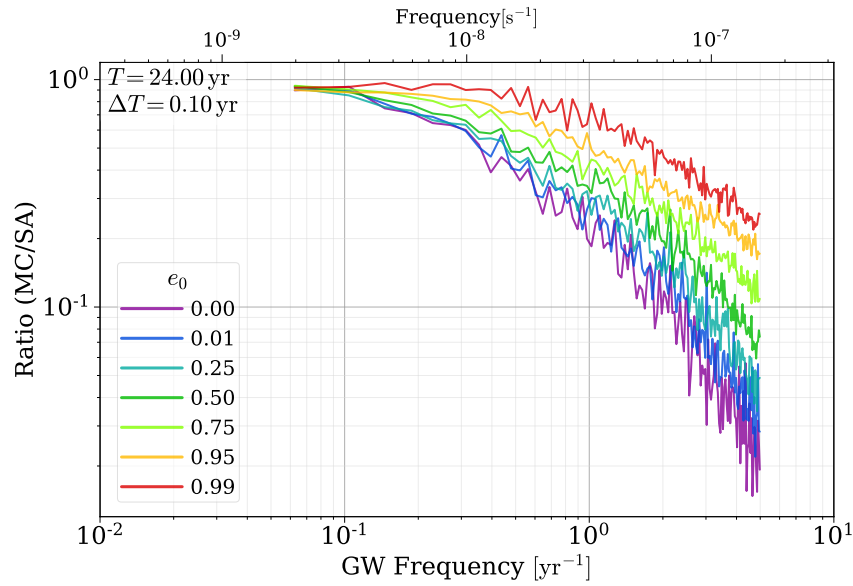
**Figure 2.12:** Upper-panel: distribution of stellar densities at 10 pc for Illustris galaxies hosting MBH binaries. For galaxies in which 10 pc is unresolved, the density is calculated from power-law fits to the innermost (resolved) regions. The dashed grey lines show the entire population of binary host galaxies, while the blue lines show hosts of *heavy* ( $M > 10^9 M_\odot$ ) binaries, and red the *heavy* and *major* ( $\mu > 0.1$ ) binaries. Each population is plotted fractionally, but note that *heavy* binaries constitute  $\sim 7\%$  and *heavy & major*  $\sim 1\%$  of all binaries respectively. Vertical lines indicate the median value of each subset.

(Snyder et al. 2015; Kelley, Blecha & Hernquist 2017).

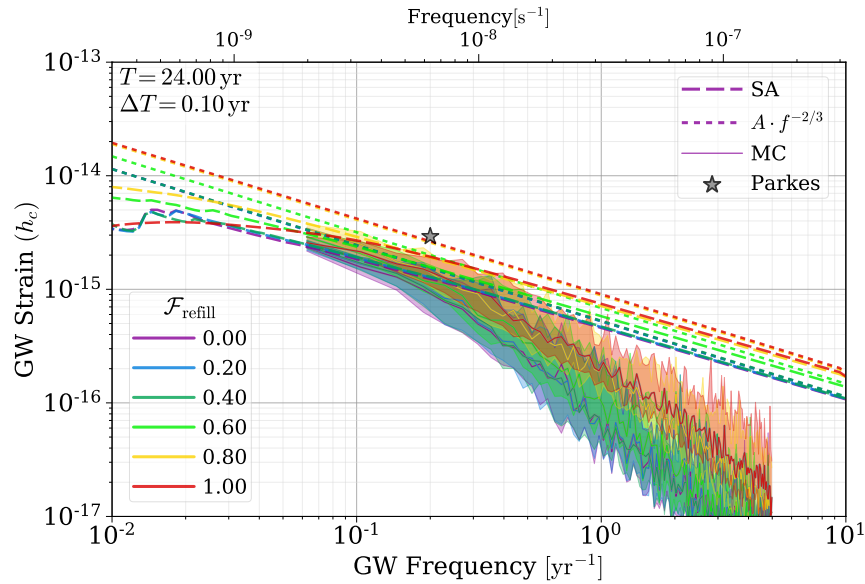
## ADDITIONAL FIGURES



**Figure 2.13:** Detection Probability (DP) for purely power-law GWB spectra exploring different intrinsic GWB amplitudes over different observation durations for each PTA. The left column shows the expanded (+) configurations where new pulsars are added each year, while the right column shows the static configurations with only the current number of pulsars. The IPTA<sub>0</sub> is the official specification for the International pulsar timing array, while the IPTA' is a more optimistic, manual combination of the specifications for each of the three individual PTA (see §2.2.4). The IPTA<sub>0</sub> is the official specification for the International pulsar timing array, while the IPTA' is a more optimistic, manual combination of the specifications for each of the three individual PTA (see §2.2.4). The horizontal, dashed grey lines show the GWB amplitude from our fiducial model:  $A_{yr^{-1}} = 0.6 \times 10^{-15}$ , and the vertical, dashed orange lines show the time at which each configuration reaches DP = 50% (short-dashes) and DP = 95% (long-dashes) for the fiducial amplitude. For a power-law spectrum at the fiducial amplitude, we expect IPTA'+ to reach 50% & 95% DP in about 3 & 8 years ( $\sim 2020$  &  $\sim 2025$ ) respectively. Without expansion, the IPTA' reached 50% DP in about 6 years ( $\sim 2023$ ), and does not reach 95% DP within 20 years.

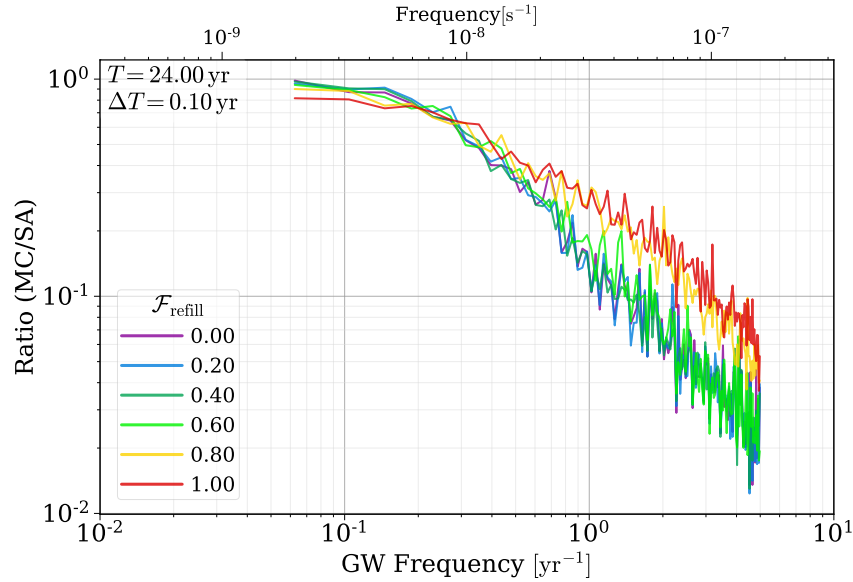


**Figure 2.14:** Ratio of the (median) MC-calculated GWB to that of the SA calculation. Binaries at higher eccentricities contribute more GW energy to higher-harmonics above their orbital frequency. This causes the number of sources contributing at higher-frequencies to increase with increasing eccentricity, decreasing the effects of MBHB quantization.



**Figure 2.15:** Monte-Carlo (MC) calculated GWB for a variety of different LC refilling fractions ( $\mathcal{F}_{\text{refill}}$ ), with median lines (solid) and one-sigma contours shown. Both the full (dashed) and power-law (dotted) SA models are also plotted. More efficient LC refilling means more binaries coalesce, causing the GWB amplitude to increase. An always full LC causes increased attenuation at lower frequencies: apparent at  $f \lesssim 0.1 \text{ yr}^{-1}$ . The steepening of the spectral index at higher frequencies due to finite-number effects is also apparent, but for  $\mathcal{F}_{\text{refill}} \geq 0.8$ , the effect is somewhat remediated.





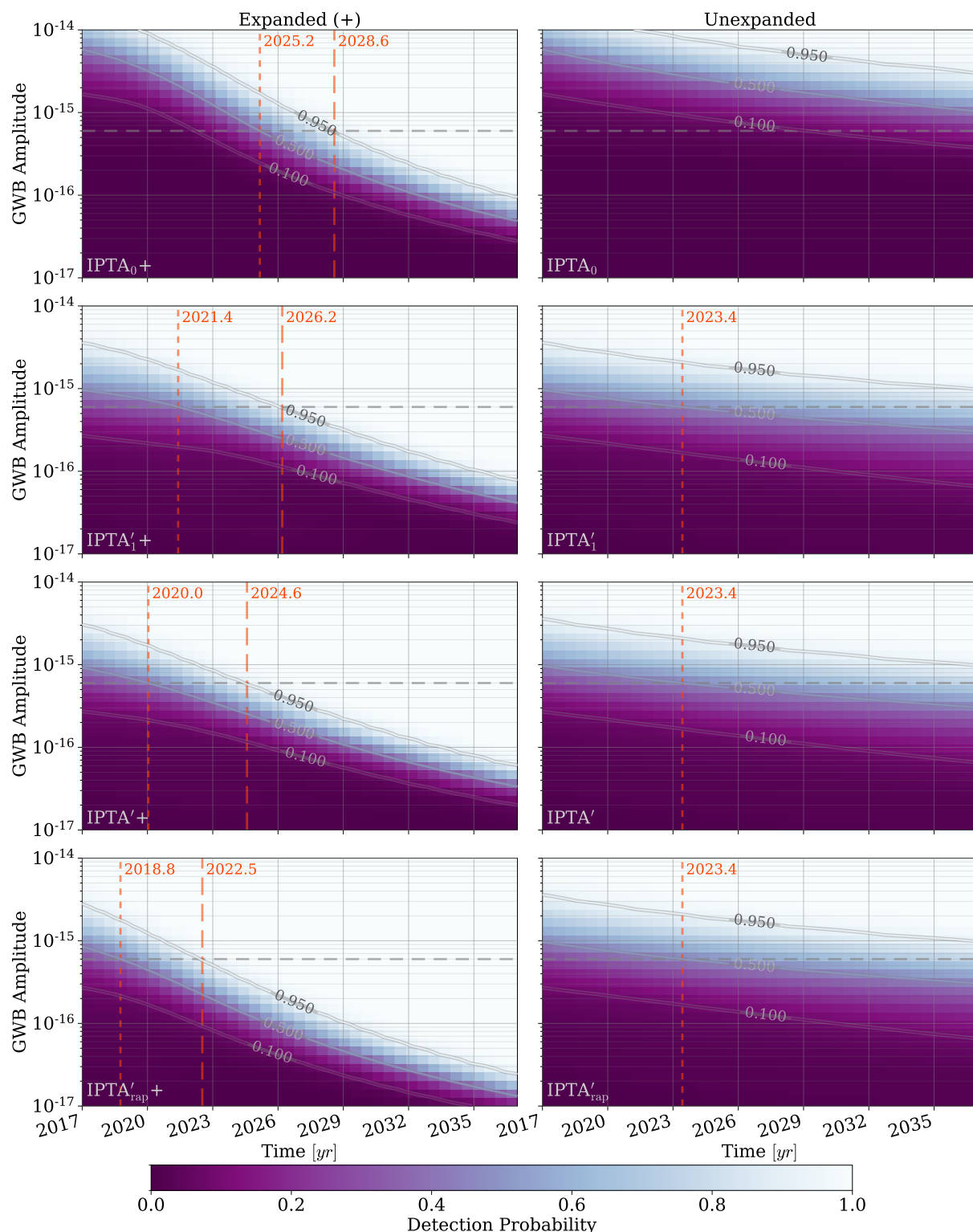
**Figure 2.16:** Ratio of the (median) MC-calculated GWB to that of the SA calculation, for zero-eccentricity and a variety of LC refilling fractions ( $\mathcal{F}_{\text{refill}}$ ). Finite-number effects, from few binaries in each bin, cause the strong deviation between Semi-Analytic (SA) and MC calculations at higher frequencies. This effect is somewhat alleviated by effective LC refilling ( $\mathcal{F}_{\text{refill}} \geq 0.8$ ) where the total number of coalescing binaries is increased.

#### INTERNATIONAL PTA MODELS AND TIME-TO-DETECTION SENSITIVITIES

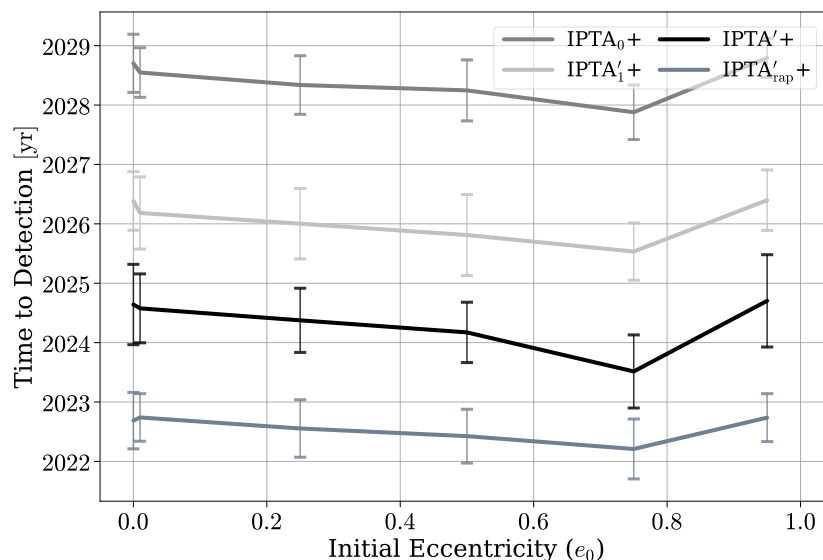
In this section we discuss different aspects of models for the International PTA. The IPTA<sub>0</sub> model is based on the official, public IPTA data release (Verbiest et al. 2016). Throughout this paper we have also focused on the IPTA' model (discussed in §2.2.4), which is a manual combination of the public data sets from each of the three individual PTA. Table 2.1 shows a summary of the differences between these primary PTA models. The time at which PTA will make detections depends sensitively on how they expand: how rapidly they add new pulsars to the arrays, and what the timing parameters of those pulsars are. We use 'expanded' PTA models (denoted with a '+') to account for this growth. The IPTA'+ model gradually increases the rate of expansion from 2011 to 2015, to account for the staggered ends of the individual PTA data sets. Initially the IPTA'+ expands by 2 pulsars per year (after 2011), and finally by 6 per year (after 2015). Here, we also introduce a IPTA'<sub>1</sub>+ model which does not expand at all until after 2015, at which point it adds 6 pulsars per year. Finally, we also show a model which uses the same expansion schedule as IPTA'+, but in which the pulsars added have a rapid cadence of 2 days, called IPTA'<sub>rap</sub>+

Figure 2.17 shows detection probability for different purely power-law GWB amplitudes for the different IPTA models. The IPTA<sub>0</sub> and IPTA'<sub>1</sub> differ in the overall pulsar parameters (noise, cadence, etc). In the expanded cases, the differences in IPTA<sub>0</sub>+ and IPTA'<sub>1</sub>+ models lead to differences of 4 & 2.5 years to reach 50% & 95% DP respectively for our fiducial amplitude of  $A_{\text{yr}^{-1}} = 0.6 \times 10^{-15}$ . In the unexpanded models the difference is even more pronounced where by 2037 the IPTA<sub>0</sub> hardly reaches 50% DP for an amplitude of  $A_{\text{yr}^{-1}} = 10^{-15}$ , while IPTA'<sub>1</sub> reaches 50% DP for  $A_{\text{yr}^{-1}} = 0.6 \times 10^{-15}$  in  $\sim 2023$ .

The IPTA'<sub>1</sub>+, IPTA'+, and IPTA'<sub>rap</sub>+ models differ in only their expansion specifications so their detection probabilities in the unexpanded configurations are identical. The expanded versions however differ notably. IPTA'<sub>1</sub>+ vs. IPTA'+ (expanding after 2015 vs. gradually increasing expansion starting in 2011) differ in time to detection by  $\sim 1.5$  yr for both 50% and 95% DP (again at  $A_{\text{yr}^{-1}} = 0.6 \times 10^{-15}$ ).



**Figure 2.17:** Detection probability for purely power-law GWB spectra of varying amplitudes versus observing time. Shown are the four different International PTA configurations discussed above, with differences in the pulsar characteristics ( $IPTA_0$  versus the other models) and different expansion specifications for adding pulsars ( $IPTA'_1+$ ,  $IPTA'+$ , and  $IPTA'_{rap}+$ ).



**Figure 2.18:** Time to detection versus initial eccentricity for the same four International PTA configurations discussed above. Differences in pulsar characteristics (most notably noise properties) and expansion prescriptions yield a 6 year range in times to detection.

Going from the IPTA'+ model to the IPTA'<sub>rap</sub>+ model (decreasing the observing cadence of added pulsars from every  $\sim 14$  days to every 2 days) further decreases the time to detection by 1 & 2 years for 50% and 95% DP.

Figure 2.18 shows time to detection (at 95% DP) versus initial eccentricity for the full GWB calculation. Overall, differences between IPTA models lead to a 6 year range of possible times-to-detection for the same GWB spectra. This highlights 1) **the importance of the red-noise characterization of pulsars**, which often disagree significantly between different PTA but for the same pulsar; 2) that the expansion prescriptions we are using are *ad hoc*, and **updates from the individual PTA and especially the IPTA on their current data sets are very important** moving ahead; and 3) that **higher cadence observations (i.e. more telescope time) will make a noticeable improvement in time to detection** (and likely how quickly SNR will grow after detection) even without consider the benefits to noise characterization.

# SINGLE SOURCES IN THE LOW-FREQUENCY GRAVITATIONAL WAVE SKY: PROPERTIES AND TIME TO DETECTION BY PULSAR TIMING ARRAYS

*This thesis chapter originally appeared in the literature as*

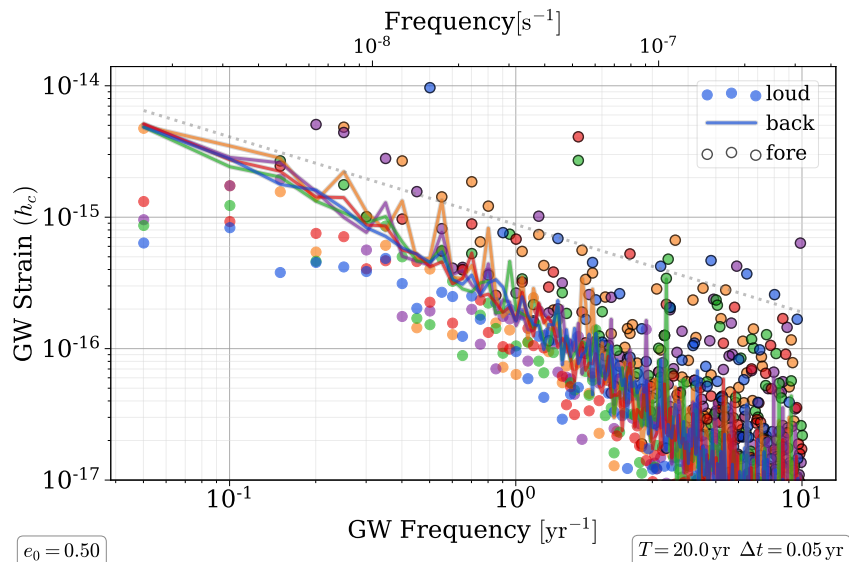
Luke Zoltan Kelley, Laura Blecha, Lars Hernquist, Alberto Sesana, Stephen R. Taylor  
MNRAS, *in press*. arXiv:1711.00075

## ABSTRACT

WE calculate the properties, occurrence rates and detection prospects of individually resolvable ‘single sources’ in the low frequency gravitational wave (GW) spectrum. Our simulations use the population of galaxies and massive black hole binaries from the Illustris cosmological hydrodynamic simulations, coupled to comprehensive semi-analytic models of the binary merger process. Using mock pulsar timing arrays (PTA) with, for the first time, varying red-noise models, we calculate plausible detection prospects for GW single sources and the stochastic GW background (GWB). Contrary to previous results, we find that single sources are at least as detectable as the GW background. Using mock PTA, we find that these ‘foreground’ sources (also ‘deterministic’/‘continuous’) are likely to be detected with  $\sim 20$  yr total observing baselines. Detection prospects, and indeed the overall properties of single sources, are only moderately sensitive to binary evolution parameters—namely eccentricity & environmental coupling, which can lead to differences of  $\sim 5$  yr in times to detection. Red noise has a stronger effect, roughly doubling the time to detection of the foreground between a white-noise only model ( $\sim 10 - 15$  yr) and severe red noise ( $\sim 20 - 30$  yr). The effect of red noise on the GWB is even stronger, suggesting that single source detections may be more robust. We find that typical signal-to-noise ratios for the foreground peak near  $f = 0.1 \text{ yr}^{-1}$ , and are much less sensitive to the continued addition of new pulsars to PTA.

## 3.1 INTRODUCTION

Pulsar timing arrays (PTA) can measure gravitational waves (GW) by precisely measuring the advance and delay in pulses from galactic millisecond pulsars (Hellings & Downs 1983). The absence of deviations in the timing of a single pulsar can be used to place constraints on the presence of GW signals (Estabrook & Wahlquist 1975; Sazhin 1978; Detweiler 1979; Bertotti, Carr & Rees 1983; Blandford, Romani & Narayan 1984) while the cross-correlation of timing data from an array of pulsars can be used to



**Figure 3.1:** Five realizations of the low-frequency GW sky are shown from our models based on black holes from the Illustris simulations. The GW spectrum is separated into the loudest individual sources per frequency-bin (circles) and the remaining ‘background’ of all other systems (lines). Single sources which are louder than the background are highlighted (black circles), and constitute the GW ‘foreground’. The power-law GWB spectrum (Eq. 3.1), assuming a continuum of sources evolving purely due to GW emission is shown with the dotted grey line.

directly measure metric deviations (Foster & Backer 1990; Jenet et al. 2005; Yardley et al. 2011; Demorest et al. 2013). The expected sources of detectable GW in the low-frequency PTA regime ( $f \sim 1 \text{ yr}^{-1} \sim 10 \text{ nHz}$ ) are massive black hole binaries (MBHB;  $M_{\text{tot}} \sim 10^6 - 10^{10} M_{\odot}$ ) in stable orbits, typically millions of years before coalescence (Rajagopal & Romani 1995; Wyithe & Loeb 2003; Jaffe & Backer 2003; Sesana et al. 2004; Enoki et al. 2004).

Three independent PTA are in operation: the European (EPTA, Kramer & Champion 2013; Desvignes et al. 2016), NANOGrav (McLaughlin 2013; The NANOGrav Collaboration et al. 2015), and Parkes (PPTA, Manchester et al. 2013; Reardon et al. 2016). The International PTA (IPTA, Hobbs et al. 2010; Verbiest et al. 2016) is a collaboration between all three which uses their combined data to boost sensitivity. Prospects for GW detection by PTA depend sensitively on the continued discovery of additional, low-noise, millisecond pulsars to incorporate into the networks (e.g. Taylor et al. 2016b). Additionally, as observing baselines on existing pulsars increase, the presence of red-noise in pulsar timing residuals can have ever increasing and more dominant effects. The source of red-noise, or even the relative contribution of astrophysical versus instrumental origins, is unclear (see, e.g. Caballero et al. 2016).

The low-frequency GW sky can be classified in terms of a stochastic GW Background (GWB)—the superposition of many unresolved GW sources; and deterministic GW sources, where individual binaries are resolvable above the background. We refer to these resolvable single sources as the GW Foreground\* (GWF). Figure 3.1 shows five randomly selected realizations of GW signals from our models, decomposed into the loudest single-source in each frequency bin (circles) and the background (all other sources combined; lines). ‘Foreground’ sources with strains larger than that of the background are highlighted in black.

\*Note that the term ‘foreground’ is also sometimes used in the literature to refer to general astrophysical sources of GW, at times even the GWB. Resolvable single sources are also referred to as ‘continuous’ and ‘deterministic’ sources.

At low frequencies, the GWB characteristic-strain tends to follow a power-law (Phinney 2001),

$$h_c \propto A_{\text{yr}^{-1}} \left( \frac{f}{1 \text{ yr}^{-1}} \right)^{-2/3}, \quad (3.1)$$

with an amplitude, typically referenced at a frequency  $f = 1 \text{ yr}^{-1}$ , on the order of  $A_{\text{yr}^{-1}} \sim 10^{-15}$  (Wyithe & Loeb 2003). The power law model is based on the assumption of a continuum of sources, evolving purely due to GW emission and over an indefinite frequency range. The  $-2/3$  spectral index emerges based on the rate at which the binary separation ‘ $a$ ’ shrinks (‘hardens’) due to GW radiation (see, e.g., the derivation in Sesana, Vecchio & Colacino 2008). At small separations the residence time,  $a/(da/dt)$ , decreases enough that the average number of binaries emitting in a given frequency bin reaches order unity, causing the spectrum to steepen and falloff from the power-law prediction (Sesana, Vecchio & Colacino 2008).

While neither class of low-frequency GW signal has been observed, the most recent upper-limits on a GWB (Lentati et al. 2015; Arzoumanian et al. 2016; Shannon et al. 2015; Verbiest et al. 2016) are now astrophysically constraining, if not surprising (e.g. Kelley, Blecha & Hernquist 2017; Middleton et al. 2017). Additionally, some groups have even begun to constrain the presence of individual MBHB in nearby galaxies (e.g. Babak et al. 2016; Schutz & Ma 2016).

Recent studies predict that the IPTA could detect the GWB by about 2030, if fiducial parameters are reasonable (e.g. Taylor et al. 2016b; Kelley et al. 2017). It is often stated that the GWB is expected to be detected before the foreground. However, only a few papers exist with quantitative predictions for single-source rates (Sesana, Vecchio & Volonteri 2009; Ravi et al. 2014)\*, and only one which includes calculations of detection probabilities (Rosado, Sesana & Gair 2015). Sesana, Vecchio & Volonteri (2009) use dark matter halos and merger rates from the Millennium simulations combined with a variety of observational BH–galaxy scaling relations to calculate resolvable MBHB strains. They predict at least one system with residuals between  $\sim 5 - 50 \text{ ns}$  after observations  $T = 5 \text{ yr}$  duration, with resolvable sources tending to come from redshift  $z \sim 0.3 - 1$ , and chirp-masses  $\log(\mathcal{M}/M_\odot) \sim 8.5 - 9.5$ . Ravi et al. (2014) use observationally determined galaxy merger rates and MBH–galaxy scaling relations to construct semi-analytic MBHB systems. At  $f_{\text{GW}} = 1 \text{ yr}^{-1}$  (with  $T = 10 \text{ yr}$ ), they expect roughly one source with a strain above  $10^{-16}$ , and a probability of  $10^{-2} - 10^{-1}$  for strains above  $10^{-15}$ . Using an IPTA-like model, Rosado, Sesana & Gair (2015) find a single-source detection probability of  $\sim 10\%$  ( $\sim 20\%$ ) after 20 yr (30 yr) of observations, and overall a  $\sim 5 - 25\%$  chance that a single-source will be detected before the GWB.

Single sources offer a largely independent window into MBHB populations and their evolution. While the amplitude of the GWB, for instance, suffers from degeneracies between the distribution of MBH masses and their coupling to environmental hardening mechanisms, the observation of deterministic sources could break that degeneracy and possibly demonstrate MBHB orbital evolution in real time. Additionally, unlike the GWB, foreground sources immediately offer the prospect of observing electromagnetic counterparts. Numerous existing surveys have already identified candidate MBHB systems based on spectroscopic and photometric techniques, for example searching for periodic variability in the CRTS (Charisi et al. 2016) and PTF surveys (Graham et al. 2015a). While there are signs that a large fraction of these candidates could be false positives (Sesana et al. 2017a), they are still promising for the possibility of multi-messenger observations.

\*Roebber et al. (2016) do not discuss single-sources *per se*, but much of the analysis is relevant, and their Fig. 2 of a high-resolution realization of GW sources, is very informative.

## 3.2 METHODS

### 3.2.1 MBH BINARY POPULATION AND EVOLUTION

In this section we summarize the key aspects of our methodology which are described in detail in [Kelley, Blecha & Hernquist \(2017\)](#) & [Kelley et al. \(2017\)](#). We use the galaxies and black hole populations obtained from the Illustris simulations, which coevolve hydrodynamic gas cells along with star, dark matter, and black hole particles over cosmic time ([Vogelsberger et al. 2014b](#); [Genel et al. 2014](#); [Torrey et al. 2014](#); [Nelson et al. 2015](#)). Once a galaxy grows to a halo mass of  $\sim 7 \times 10^{10} M_{\odot}$  ( $M_{\text{star}} \sim 10^9 M_{\odot}$ ) it is given a MBH with a seed-mass of  $10^5 M_{\odot}$ , which then accretes matter from the neighboring gas cells as the galaxy evolves ([Vogelsberger et al. 2013](#); [Sijacki et al. 2015](#)). After a galaxy merger, if two MBH particles come within a gravitational smoothing length of one another (typically  $\sim \text{kpc}$ ), they are manually ‘merged’ into a single MBH with their combined masses.

We identify these *pseudo*-mergers in Illustris and further evolve the constituent MBH in semi-analytic, post-processing simulations to model the effects of the small scale (sub-kiloparsec) ‘environmental’ processes which mediate the true, astrophysical merger process. To do this, we calculate density and velocity profiles from each MBHB host galaxy, and use them to calculate hardening rates from dynamical friction, stellar scattering, viscous drag (from a circumbinary disk), and gravitational wave emission. Eccentric binary evolution can also be included, in which the eccentricity is enhanced by stellar scatterings and decreased by gravitational wave emission (discussed in [Kelley et al. 2017](#)). When binaries are initialized following their pseudo-mergers in Illustris, they are given a uniform, fixed value of eccentricity ( $e_0$ ), which, along with the binary separation, is then numerically integrated in time until each system either coalesces or reaches redshift  $z = 0$ . Due to the scarcity of MBHB observations, there are virtually no constraints on their eccentricity distribution. We thus explore a wide range of initial eccentricities,  $e_0 = [0.0, 0.01, 0.25, 0.50, 0.75, 0.95, 0.99]$ . For circular evolution models, we use a stellar scattering prescription based on [Magorrian & Tremaine \(1999\)](#), in which we can vary the effectiveness of scattering based on a parameter\*  $\mathcal{F}_{\text{refill}} \in [0.0, 1.0]$  — which acts as a proxy for the overall degree of environmental coupling (see [Kelley, Blecha & Hernquist 2017](#)).

By associating the presence of each binary in the simulation volume ( $V_{\text{ill}} = 80 \text{Mpc}^3$ , at  $z = 0$ ) with the result of a Poisson process, we can calculate GW spectra from an arbitrary number of observational ‘realizations’ by re-drawing from the appropriate distribution and weighting each binary accordingly. Specifically, we define a *representative volume factor* for binary  $i$  at a time-step  $j$ ,

$$\Lambda_{ij} = \frac{1}{V_{\text{ill}}} \frac{dV_c(z_{ij})}{dz_{ij}} \Delta z_{ij}, \quad (3.2)$$

where the comoving volume element is,

$$dV_c(z) = 4\pi (1+z)^2 \frac{c}{H(z)} d_c^2(z) dz. \quad (3.3)$$

Here,  $H(z)$  is the Hubble constant at redshift  $z$  (and corresponding comoving distance  $d_c$ ),  $c$  is the speed of light, and  $\Delta z_{ij}$  is the redshift step-size for this binary and time step.  $\Lambda_{ij}$  is the expectation value for the number of *astrophysical* binaries in the past light-cone between redshifts  $z_{ij}$  and  $z_{ij} + \Delta z_{ij}$ , corresponding to each *simulated* binary & time-step. To construct an alternate realization, we can draw from the Poisson distribution,  $\mathcal{P}(\Lambda_{ij})$ , for each binary and step. For example, the characteristic GW strain spec-

\*This ‘refilling’ parameter interpolates between a sparsely filled loss-cone (the region of stars in parameter space able to interact with the binary), and a full one.

trum from all binaries is,

$$h_c^2(f) = \sum_{ij} \mathcal{P}(\Lambda_{ij}) \sum_{n=1}^{\infty} \left[ \frac{f}{\Delta f} h_s^2(f_r) \left( \frac{2}{n} \right)^2 g(n, e) \right]_{f_r=f(1+z)/n}. \quad (3.4)$$

The spectrum  $h_c^2$  is over all observed frequencies  $f$  with bins of width  $\Delta f$ , usually determined by the total temporal baseline of observations (e.g. by PTA). Equation 3.4 takes into account that an eccentricity  $e$  leads to a redistribution of GW energy to each harmonic  $n$  of the rest-frame orbital frequency  $f_r$ . The amount of power observed at  $f = f_r n/(1+z)$  is given by the power distribution function  $g(n, e)$  (see, Eqs. A1 in [Kelley et al. 2017](#)). For a chirp-mass  $\mathcal{M} = (M_1 M_2)^{3/5} / (M_1 + M_2)^{1/5}$ , the sky- and polarization- averaged GW-strain from a binary with zero eccentricity is (e.g. [Sesana et al. 2004](#)),

$$h_s(f_r) = \frac{8}{10^{1/2}} \frac{(G\mathcal{M})^{5/3}}{c^4 d_c} (2\pi f_r)^{2/3}, \quad (3.5)$$

which, for a circular binary, is emitted entirely at the  $n = 2$  harmonic. The *characteristic* strain for a particular source, which takes into account the number of cycles viewed in-band, is,

$$h_{c,s}^2(f_r) = h_s^2(f_r) \left( \frac{f}{\Delta f} \right). \quad (3.6)$$

Alternatively, the observed timing residual can be calculated directly\* as ([Sesana, Vecchio & Volonteri 2009](#)),

$$\delta t = \frac{10^{1/2}}{15} h_s(f) \left( \frac{T}{f} \right)^{1/2}. \quad (3.7)$$

### 3.2.2 DETECTION STATISTICS AND PULSAR TIMING ARRAY MODELS

In our analysis we only consider the single loudest sources in each frequency bin as candidates for the foreground. [Ravi et al. \(2012\)](#) point out that PTA can resolve spatially in addition to chromatically, allowing multiple loud sources to be simultaneously extracted from the same bin. [Boyle & Pen \(2012\)](#) and [Babak & Sesana \(2012\)](#) demonstrate that this is possible if there are roughly six or more pulsars equally dominating PTA sensitivity. The uncertainty introduced by neglecting this effect is small compared to those of the models being used, and additionally, based on our results, it is rare for multiple single-sources to each produce comparable strains while also resounding over the GWB.

Methods for the detection of single sources by existing PTA have been rapidly developed in the last decade (e.g. [Corbin & Cornish 2010](#); [Lee et al. 2011](#); [Boyle & Pen 2012](#); [Babak & Sesana 2012](#); [Ellis, Siemens & Creighton 2012](#); [Ellis 2013](#); [Taylor, Ellis & Gair 2014](#); [Zhu et al. 2015, 2016](#); [Taylor et al. 2016a](#)). For detection statistics with mock PTA, we use the methods presented in [Rosado, Sesana & Gair \(2015, hereafter, RSG15\)](#). The statistics for the background, based on cross-correlations between pulsars, were also used in [Kelley et al. \(2017\)](#). In this study, however, we explore the distinction between a *true* background with the loudest single-sources per bin removed (‘back’) at the same time as *all* GW sources included (‘both’).

The formalism for single-sources (GWF), based on excess power recovery<sup>†</sup>, requires all pulsars to have the same frequency sampling determined by the observing duration  $T$  (which we vary) and cadence  $\delta t$  (fixed to  $\delta t = 0.05$  yr). We use Eq. 35 & 36<sup>‡</sup> ([RSG15](#)) to calculate the signal-to-noise ratio (SNR) for each pulsar and each frequency bin, given a GW spectrum and PTA configuration. Using Eq. 32 ([Ibid.](#))

\*  $\delta t \propto h_s \tau_{\text{gw}} N^{1/2}$ , where the observed GW period  $\tau_{\text{gw}} = 1/f$ , and the number of observed cycles  $N = fT$ .

<sup>†</sup>and is thus not an optimal statistic

<sup>‡</sup>Compared to [RSG15](#), we define the GW phase after a time  $T$  to be  $\Phi_T = 2\pi fT + \Phi_0$ , for an initial phase  $\Phi_0$ . We find a slightly different expression for the signal-to-noise after integrating their Eq. 36 than they present in Eq. 46, but the differences are negligible once incorporated into a full analysis.



| name     | $h_c^N(f = 0.1 \text{ yr})$             | $\sigma_{\text{WN}}$ | $A_{\text{RN}}$     | $\gamma_{\text{RN}}$ |
|----------|---|----------------------|---------------------|----------------------|
| <b>a</b> | <b><math>1.0 \times 10^{-15}</math></b> | $0.3 \mu\text{S}$    | -                   | -                    |
| b        | $1.8 \times 10^{-15}$                   | $0.3 \mu\text{S}$    | $4 \times 10^{-15}$ | -3.0                 |
| c        | $6.7 \times 10^{-15}$                   | $0.3 \mu\text{S}$    | $1 \times 10^{-15}$ | -4.5                 |
| <b>d</b> | <b><math>1.8 \times 10^{-14}</math></b> | $3.0 \mu\text{S}$    | $4 \times 10^{-15}$ | -3.5                 |
| <b>e</b> | <b><math>4.6 \times 10^{-14}</math></b> | $3.0 \mu\text{S}$    | $2 \times 10^{-13}$ | -1.5                 |
| f        | $2.0 \times 10^{-13}$                   | $0.3 \mu\text{S}$    | $2 \times 10^{-13}$ | -3.0                 |

**Table 3.1:** Parameters of the noise models used in our mock PTA. The 2nd column gives the total noise, in units of characteristic strain, at  $f = 0.1 \text{ yr}^{-1}$ . The remaining columns are the white-noise RMS  $\sigma_{\text{WN}}$ , and the red-noise amplitude & spectral index,  $A_{\text{RN}}$  &  $\gamma_{\text{RN}}$  (see: Eq. 3.8). The parameters for these models were chosen manually while trying to accurately represent the properties of observed pulsars (see: Fig. 3.12). Much of our analysis focuses on models ‘a’, ‘d’ & ‘e’, which are highlighted.

the SNR are converted to bin/pulsar detection probabilities (DP), and Eq. 33 (Ibid.) finally converts to an overall DP for the PTA to detect at least one GWF source. The single-source detection statistics we use are not designed for eccentric systems, where the signal from an individual source can be spread over multiple frequency bins. When this happens, we are effectively treating the portion of the signal in each bin as independent, which is obviously sub-optimal. As most of the binaries contributing to the GW signals have fairly low eccentricity (see, e.g. Fig. 3.8; and Kelley et al. 2017), this should only have a minor effect.

To calculate plausible detection probabilities, we use mock PTA with a variety of noise models. We characterize the noise with three parameters, a white-noise amplitude  $\sigma_{\text{WN}}$ , and a red-noise amplitude & spectral index  $A_{\text{RN}}$  &  $\gamma_{\text{RN}}$  for a power-spectrum,

$$S_{\text{RN}} = \frac{A_{\text{RN}}^2}{12\pi^2} \left( \frac{f}{f_{\text{ref}}} \right)^{\gamma_{\text{RN}}} f_{\text{ref}}^{-3}. \quad (3.8)$$

Defined in this way,  $A_{\text{RN}}$  corresponds to the equivalent strain at the reference frequency  $f_{\text{ref}}$ , which is always set to  $1 \text{ yr}^{-1}$  in our analysis. All pulsars in a given array use the same noise parameters.

Noise models were selected to cover a parameter space comparable to that measured by PTA. The noise characterization between different PTA can vary significantly, at times being inconsistent. For example, the red noise of J1713+0747 is characterized by an amplitude and spectral index:  $A_{\text{RN}} = 2 \times 10^{-15}$  &  $\gamma_{\text{RN}} = -4.8$  by the EPTA (Caballero et al. 2016), and  $A_{\text{RN}} = 3.5 \times 10^{-14}$  &  $\gamma_{\text{RN}} = -2.0$  by Parkes (Reardon et al. 2016). Similarly, J1910+1256 is given  $A_{\text{RN}} = 2.8 \times 10^{-13}$  &  $\gamma_{\text{RN}} = -1.9$  by NANOGrav (The NANOGrav Collaboration et al. 2015), but  $A_{\text{RN}} = 2.9 \times 10^{-15}$  &  $\gamma_{\text{RN}} = -5.9$  in the IPTA data release (Verbiest et al. 2016). The nature and characterization of pulsar red-noise would thus seem to be quite uncertain. The parameters of the six noise models we use are given in Table 3.1, and are plotted against pulsars from each PTA in Fig. 3.12. Much of our analysis focuses on noise models ‘a’, ‘d’ & ‘e’, which we consider as qualitatively *optimistic*, *moderate*, & *pessimistic* respectively.

The properties of the GWF depend not only on the distribution of the loudest sources, but also the frequency bin-width, and the distribution of all systems in that bin. These effects are taken into account in the detection statistic by including the background in the ‘noise’ term, when calculating an SNR. A sample calculation of the strain, SNR, and detection probability (DP) for the GWF are shown in Fig. 3.15, for noise-model ‘d’, at a number of sample frequencies. While the single-source strains and resulting SNR are distributed around a well-defined peak, the distributions of DP end up being fairly flat because of their strong sensitivity to SNR. These relatively flat DP distributions lead to large variations between realizations of the foreground. The error bars included in our figures should thus be kept in mind. The standard

deviations in DP are fairly insensitive to increases in the number of realizations we consider, implying that the size of our underlying MBHB population may be the limiting factor (discussed further in §3.4.1).

To analyze the properties of the GWF, we define its sources as those which contribute at least a fraction  $\lambda_{\text{fore}}$  of the total GW energy in that bin, i.e.  $(h_c^{\text{fore}})^2 \gtrsim \lambda_{\text{fore}} (h_c)^2 = \lambda_{\text{fore}} [(h_c^{\text{fore}})^2 + (h_c^{\text{back}})^2]$ . In our analysis we explore different values of  $\lambda_{\text{fore}}$ , but adopt a fiducial value of  $\lambda_{\text{fore}} = 0.5$ , as used in Fig. 3.1. In practice, the level at which a single source is discernible will depend on the overall source and PTA properties, especially via the signal-to-noise ratio (SNR), i.e.  $\lambda_{\text{fore}} \propto \text{SNR}^{-1}$ . With a sufficient number of pulsars contributing comparatively to the SNR, a PTA can even discern multiple single-sources in a single frequency-bin (Babak & Sesana 2012; Boyle & Pen 2012; Petiteau et al. 2013), making  $\lambda_{\text{fore}} = 0.5$  conservative in the eventual high-SNR regime.

### 3.3 RESULTS

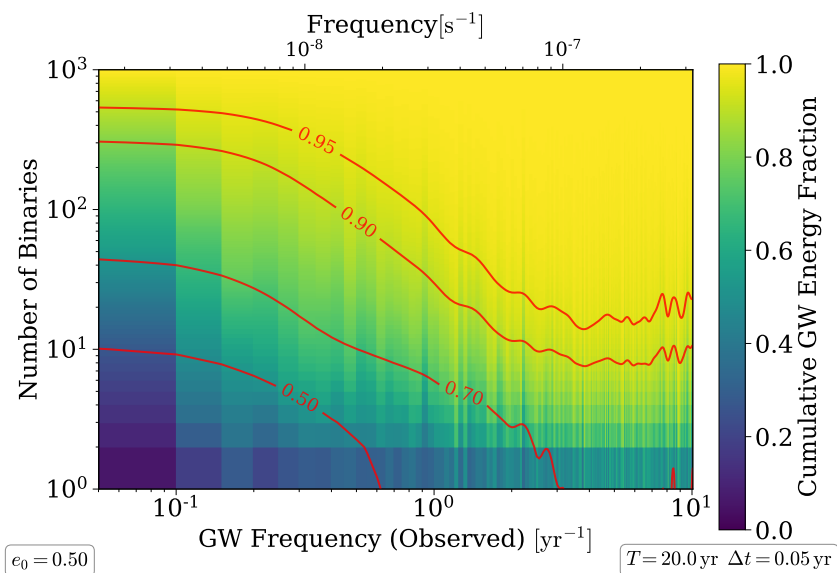
#### 3.3.1 THE STRUCTURE OF THE LOW-FREQUENCY GW SKY

Figure 3.2 shows the median, cumulative fraction of GW energy contributed by the loudest binaries in each frequency bin. The number of sources which contribute significantly to the GW energy density falls rapidly with increasing frequency in direct proportion to binary residence time. For a circular binary hardening solely due to GW emission, the residence time scales with frequency as  $\tau_{\text{gw}} \propto f^{-8/3}$ . At  $f_{\text{GW}} \sim (10 \text{ yr})^{-1}$ , the median number of binaries producing 50% of the GWB energy is  $\sim 10$ , while by  $f_{\text{GW}} \sim 1 \text{ yr}^{-1}$ , that number falls to  $\sim 1$ . These results can be compared to those of Ravi et al. (2012, Fig. 2) who find fairly consistent values, although more sources contributing at low frequencies, which implies a lower GWF rate.

Figure 3.3 shows the probability of a single source producing timing residuals and strains above a given value. At  $f_{\text{GW}} \sim 1 \text{ yr}^{-1}$ , 50% of our realizations have a MBH binary with strain above  $\sim 2 \times 10^{-16}$  or a timing residual of  $\sim 1 \text{ ns}$ . Our 10% expectations are about an order of magnitude lower than recent PTA upper limits for foreground sources:  $\sim 6 \times 10^{-15} - 10^{-14}$  at  $\sim (5 \text{ yr})^{-1}$  by the EPTA (Babak et al. 2016),  $3 \times 10^{-14}$  at  $\sim (3 \text{ yr})^{-1}$  by NANOGrav (Arzoumanian et al. 2014),  $2 \times 10^{-14}$  at  $\sim (3 \text{ yr})^{-1}$  by the PPTA (Zhu et al. 2014). These strains and timing residuals are a few times higher than those of Sesana, Vecchio & Volonteri (2009) who predict at least one source with timing residuals between  $\sim 5 - 50 \text{ ns}$  after  $T = 5 \text{ yr}$ .

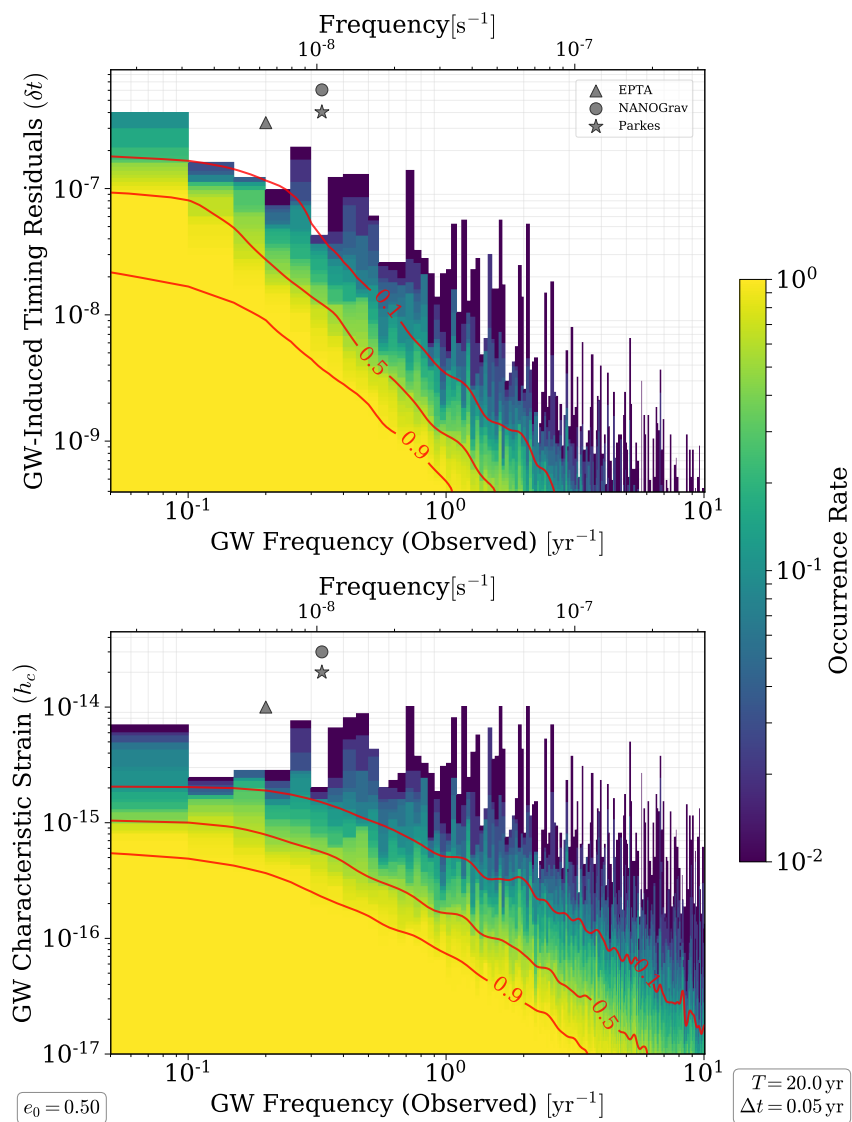
The previous figures examined the properties of all binaries. Figure 3.4 shows the fraction of realizations containing a foreground source in each frequency bin (circles) for three different foreground factors:  $\lambda_{\text{fore}} = 0.5, 0.7 \& 0.9$ . The average over fixed logarithmic frequency intervals is also shown (lines). While the typical amplitude of foreground strains decreases significantly at higher frequencies, the ever decreasing number of sources contributing at those frequencies leads to higher GWF rates.

The parameters of the binaries producing low-frequency GW signals are shown in Fig. 3.5. The properties of the loudest source in each frequency bin are plotted in red, and those of all other systems (weighted by GW energy) in blue. For comparison, the median properties of all systems (unweighted) are shown in grey. In the strain panel (upper-left), the GWB itself is shown for comparison in green as the typical strains of background sources are on the order of  $10^{-19} - 10^{-20}$ . The trends in binary parameters are driven by the convolution of hardening timescale and the number density of MBH binaries which falls strongly with total mass. Higher-mass binaries harden faster, and all systems spend less and less time at smaller separations. At low frequencies, where massive binaries are still numerous, they dominate the population and even occur at smaller distances. Overall, loud sources have much broader distributions of

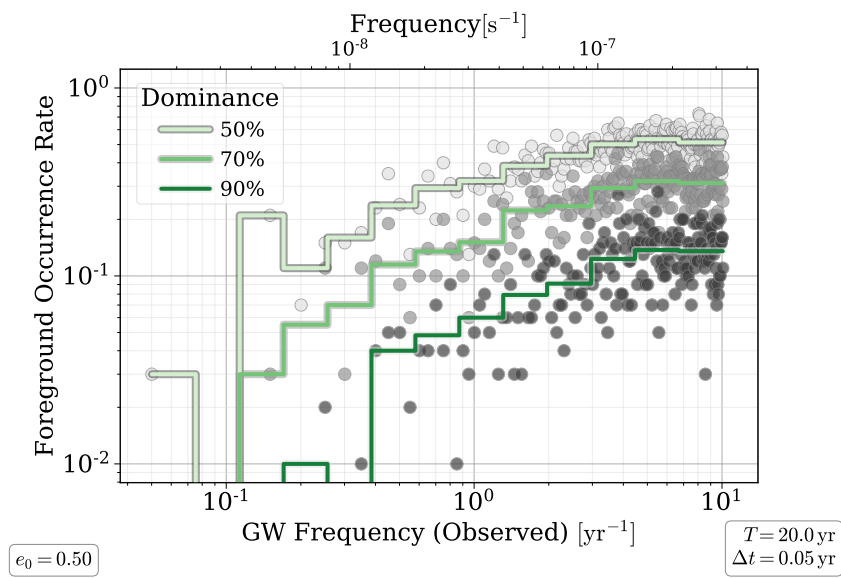


**Figure 3.2:** The cumulative fraction of GW energy produced by the loudest binaries in each frequency bin. The number of sources dominating the GW energy contribution drops rapidly with increasing frequency in correspondence with the overall number of binaries at the corresponding separations. While the relative contribution from the loudest sources increases at high frequencies, the overall amplitude of GW signals simultaneously decreases.

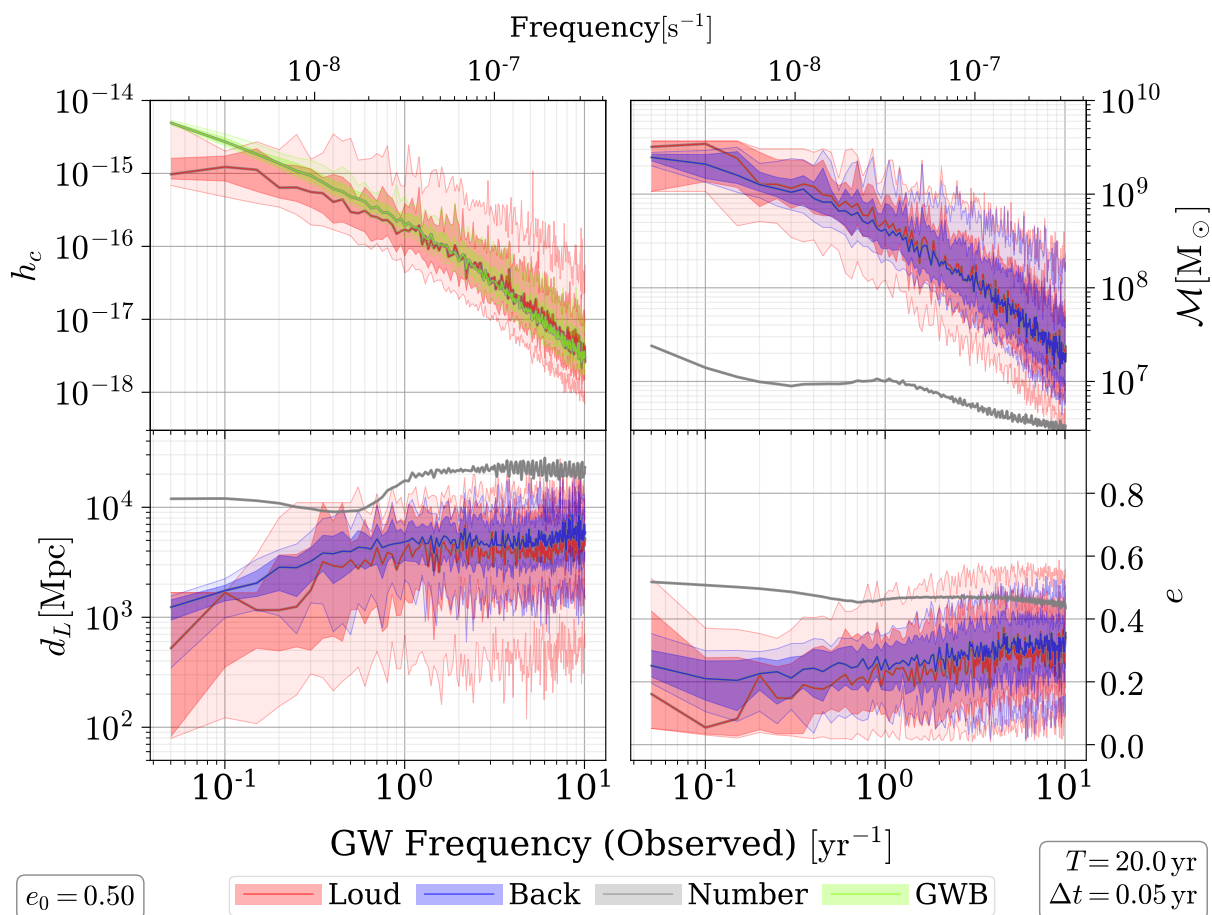
parameters, but tend to be slightly more massive, nearer, and lower eccentricity than the corresponding background sources.



**Figure 3.3:** The rate at which single-source timing residuals and strains of a given amplitude are produced in each frequency bin. Below  $f \sim 1/(10 \text{ yr})^{-1}$ , each bin typically has a source producing residuals of  $\sim 10 - 100 \text{ ns}$  ( $h_c \sim 10^{-15}$ ). The current single-source upper limits from PTA are also shown from the EPTA (Babak et al. 2016), NANOGrav (Arzoumanian et al. 2014), and PPTA (Zhu et al. 2014).



**Figure 3.4:** The fraction of realizations with a foreground source in each frequency bin (circles) and averaged over frequencies (lines). From light to dark, different criteria of foreground sources are shown:  $\lambda_{\text{fore}} = 0.5, 0.7 \& 0.9$ . Single sources which are  $\sim 10\times$  louder than the background ( $\lambda_{\text{fore}} = 0.9$ ) are generally one-tenth as common as those equal to the background ( $\lambda_{\text{fore}} = 0.5$ ), at  $f \sim 1/(5 \text{ yr})^{-1}$  occurring  $\sim 1\%$  and  $\sim 10\%$  of the time respectively.



**Figure 3.5:** Properties of low-frequency GW sources showing 1- & 2-  $\sigma$  contours over 100 realizations. Properties of the loudest source in each frequency bin (red) are compared to those of all other systems (weighted by GW-energy; blue). The median properties of all systems (unweighted) are also shown (grey). Loud sources have much broader distributions of parameters, but tend to be slightly more massive, nearer, and lower eccentricity than the corresponding background sources when the latter are weighted by their GW energy.

### 3.3.2 PARAMETRIC DEPENDENCIES

Eccentricity, by shifting the distribution of emitted GW energy versus frequency, affects the number of sources dominating GW signals. Figure 3.6 shows the number of loudest binaries contributing 0.5 of the total GW energy for each eccentricity model. The inset panel shows the trends versus eccentricity at the orange-highlighted frequencies. As eccentricity increases, more sources contribute to the GW energy at all frequencies, but the effect is strongest at low to intermediate frequencies ( $\sim 0.1 - 0.3 \text{ yr}^{-1}$ ).

Varying the stellar scattering efficiency, unlike eccentricity, has very little effect on the number of binaries contributing significant GW energy. While the effectiveness of scattering modulates the overall merger rate and total GW energy, it does not redistribute energy over frequency. This fact is echoed in the amplitudes of the loudest sources which also show virtually no dependence on environmental hardening rate. The latter is true, although to a lesser extent, with varying eccentricity.

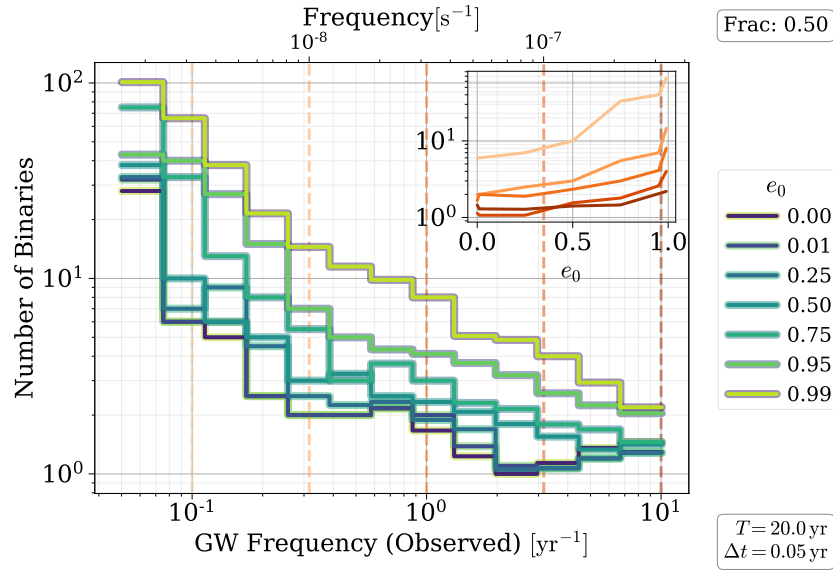
Despite the weak effect of eccentricity on the loudness of sources, their varying number noticeably alters the rate of foreground source occurrences. Figure 3.7 illustrates the rate at which a foreground source appears per frequency bin, for each eccentricity model. As eccentricity increases, the rate of foreground sources decreases. This is due primarily to two effects which occur at high eccentricities: i) more sources contribute significantly to each frequency bin (Fig. 3.6); and, ii) the GW power from each single source is spread out over a broader frequency range. Overall, Fig. 3.7 shows that initial eccentricities  $e_0 \gtrsim 0.9$  produce 2 – 10 times fewer foreground sources as  $e_0 \sim 0.0$ .

Eccentricity also slightly increases the rate at which binaries harden, which has some effect on the GWB amplitude but very little effect on the GWF occurrence rate. This can be seen in Fig. 3.13 which shows GWF rate for varying loss-cone refilling parameters ( $\mathcal{F}_{\text{refill}}$ )—a good proxy for the overall degree of environmental coupling. The increased hardening rate from  $\mathcal{F}_{\text{refill}}$  is substantially stronger than that of eccentricity (for these frequencies), and produces virtually no change to the GWF rate.

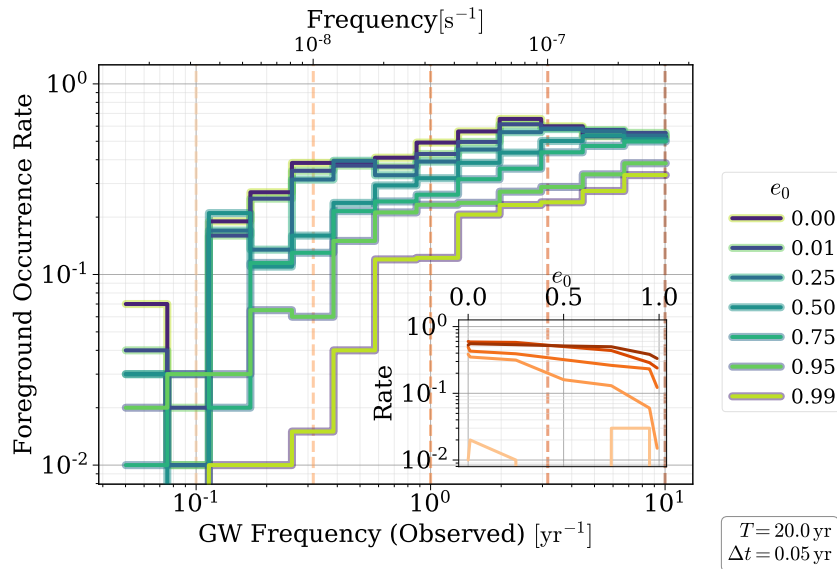
As alluded to previously, the binary parameters of sources contributing to the GWB and GWF are mostly insensitive to hardening models. Figure 3.8 shows the properties of the loudest sources in each bin (red) compared to those of the background, weighted by GW energy, (blue) as a function of initial-eccentricity, at a frequency of  $f = 0.32 \text{ yr}^{-1}$ . The distribution of unweighted properties are also plotted (grey) for comparison. In each case, a one-sigma interval is shown. In the most extreme eccentricity model,  $e_0 = 0.99$ , the distances to both loud and background sources increases drastically and the strain drops correspondingly. In this case, the eccentricity is extreme enough that the majority of the GW energy in this regime is shifted out of band. Otherwise typical properties are constant as eccentricity varies, as is the case with varying stellar scattering efficiencies.

The lower right panel in Fig. 3.8 shows the eccentricity distributions of sources at  $f = 0.32 \text{ yr}^{-1}$ . The eccentricity in the unweighted population of all binaries is nearly linear with initial eccentricity. Loud sources, and the binaries which contribute most to the background, however, have significantly damped eccentricities which are much lower in the PTA band than their initial values\*. The loudest sources tend to have lower median eccentricities by  $\sim 0.1$ , and the low-end of their 68% interval is typically half that of background sources. More massive systems circularize more rapidly, but the tendency for lower eccentricities in single sources is likely a selection bias as they are the systems which emit GW energy in a more concentrated frequency range (i.e. nearer the  $n = 2$  harmonic).

\*Kelley et al. (2017) shows in detail the evolution of eccentricity as binaries harden

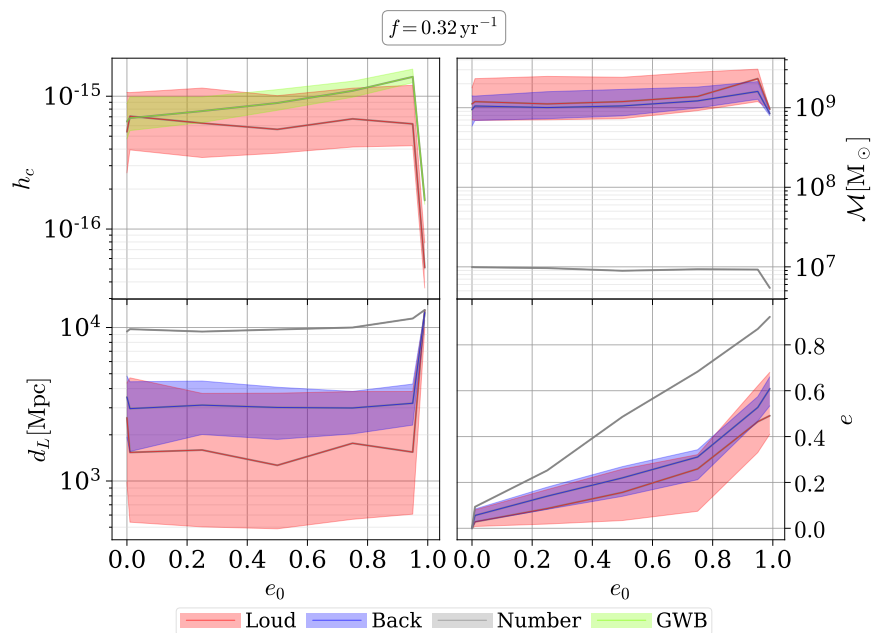


**Figure 3.6:** Number of binaries contributing 50% of the GW energy for each eccentricity model. The inset shows the trends versus eccentricity at each of the orange highlighted frequencies. As eccentricity increases, the number of contributing systems increases and the drop-off at higher frequencies becomes more gradual.



**Figure 3.7:** The occurrence rate of foreground sources ( $\lambda_{\text{fore}} = 0.5$ ) per frequency bin for each initial eccentricity model. For low-to-moderate initial eccentricities ( $e_0 \lesssim 0.5$ ) the rate plateaus to  $\sim 50\%$  at frequencies  $f \gtrsim 2 \text{ yr}^{-1}$ , and at higher eccentricities ( $e_0 \gtrsim 0.75$ ) the occurrence rate drops rapidly at lower frequencies ( $f \lesssim 1/(2 \text{ yr})^{-1}$ ). The inset panel shows the foreground occurrence rate versus eccentricity for the orange highlighted frequencies.





**Figure 3.8:** Properties of the GW sources at  $f = 0.32 \text{ yr}^{-1} = 10 \text{ nHz}$  versus initial eccentricity. Except for the highest eccentricity model ( $e_0 = 0.99$ ), the properties of the foreground are insensitive to evolutionary models. There is a slight trend for increasing strains and chirp masses with both increasing initial-eccentricity and loss-cone efficiency.

### 3.3.3 TIMES TO DETECTION

In Fig. 3.9, we show detection probability (DP) versus time for the  $e_0 = 0.5$  model, and a PTA with 60 pulsars for each noise model from Table 3.1 and Fig. 3.12. For comparison, the IPTA first data release included almost 50 pulsars, with a median observing duration of  $T \sim 11$  yr. Forecasts for expansions typically assume 6 new pulsars added to the IPTA per year (e.g. Taylor et al. 2016b), but new pulsars, with very short observing baselines, will contribute very little to DP initially. Recall, additionally, that we assume uniform noise characteristics for each pulsar in our mock arrays, while real PTA have pulsars with highly heterogenous noise characteristics which further complicates a direct comparison to our results.

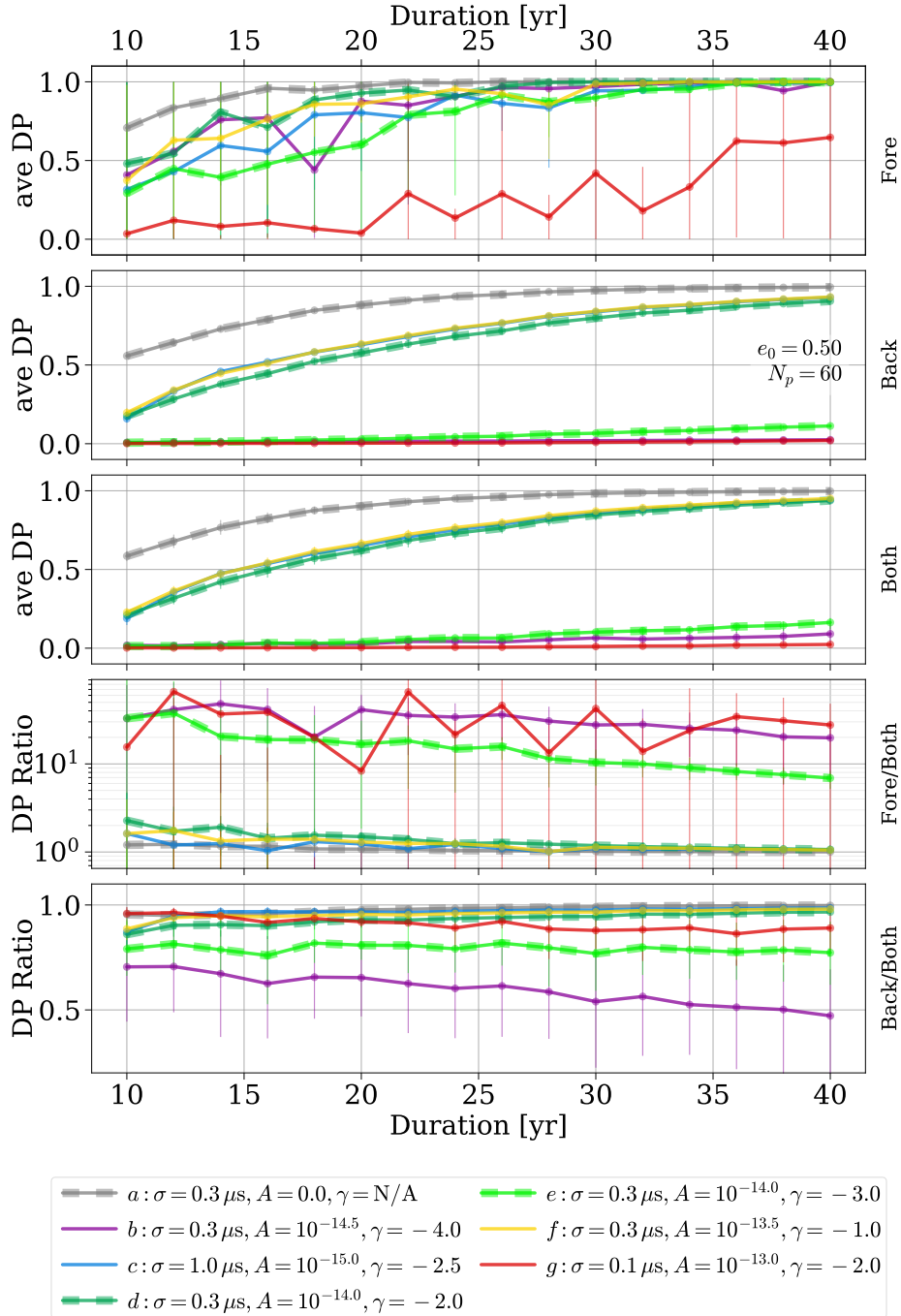
The first three panels of Fig. 3.9 show DP for the foreground (loudest source in each frequency bin) & background (all other sources in each frequency bin) and ‘both’ in which the signal is the sum of all background and foreground sources. DP for the ‘both’ model is still calculated using the detection statistics formally for a ‘background’. This is meant to examine the effect of removing the loudest sources on the background detection statistics. If we associate this mock PTA at 10 years with the current IPTA, then depending on noise model the expected DP would be anywhere between  $\sim 0.0$  for severe red noise, to  $\sim 0.7$  or  $\sim 0.5$  for the foreground and background respectively for white-noise only. Clearly, better understanding the red noise of observed pulsars is crucial to reliably forecasting low-frequency GW detections.

The last two panels show the DP ratios of foreground-to-both, and background-to-both. In all models, the GWF have uniformly higher DP than either the ‘back’ or ‘both’ signals. In the case of no red-noise (‘a’, grey), the ‘back’ and ‘both’ DP are only slightly below that of the foreground. Stronger noise, especially red noise, affects the GWB detection much more strongly than the GWF. This is not surprising as red noise, by definition, affects the lowest frequency bins most strongly, which is where the GWB is loudest, but generally below the GWF peak (see, e.g. Fig. 3.3).

With the GWB detection statistics, the noise models become highly stratified. The shallow red-noise models—‘c’, ‘d’, & ‘f’—all perform comparably with  $\sim 50\%$  lower DP than the white-noise only model at 10–15 yr. At later times the shallow models approach the white-noise only values. The steep red-noise models—‘b’, ‘e’ & ‘g’—also all perform comparably, but with near-zero DP up to  $\sim 30$  yr. The ‘both’ and ‘back’ DP typically differ by at most 50%, while ‘fore’ and ‘both’ differ by over an order of magnitude in the steep red-noise models. This suggests that the difference between GWF and GWB DP is driven largely by the nature of the detection statistic instead of simply the amount of GW power being analyzed.

To observe the GWB, the correlations between pulsars are required to distinguish the GW signal from that of noise. Because GWF sources will behave deterministically, they should be easier to distinguish. At the same time, line-like noise sources (for example due to uncertainties in planetary ephemerides), or GWF sources with few resolved periods (thus more closely resembling red noise), may complicate the identification process via single pulsars. If, instead, a correlated search is required, the recovery efficiency could become significantly lower — perhaps yielding DP which lie below that of the GWB.

Due to the similarity between numerous noise models, our additional analysis focuses on the ‘a’, ‘d’ & ‘e’ configurations. These can be considered qualitatively as optimistic, moderate & pessimistic respectively, but keep in mind that realistic PTA pulsars have heterogenous noise properties possibly making ‘a’ and ‘d’/‘e’ *overly* optimistic & pessimistic respectively (see §3.2.2). Figure 3.10 compares DP progressions for PTA with differing numbers of pulsars and these selected noise models. Because the GWB detection statistic depends on a cross-correlation between pulsars—i.e. pulsar-pairs, it is far more sensitive to the number of pulsars included in the array. Considering the ‘d’ noise-model: at  $T = 20$  yr, the difference in DP between 20 and 80 pulsars is 70 vs. 95% for the GWF, but 10 vs. 80% for the GWB. In the highly pessimistic noise-model ‘e’, 50% DP isn’t reached for the background within 40 yr, even for 100 pulsars, while



**Figure 3.9:** Detection Probability (DP) versus observing duration for mock PTA with a variety of noise models (lines). Panels show DP for the ‘Fore’ground (the loudest source in each frequency bin), the ‘Back’ground (all *other* sources in each bin), and ‘both’ in which the signal is the sum of both foreground and background sources. The ‘both’ signal is still measured using the ‘background’ detection statistics from [Rosado, Sesana & Gair \(2015\)](#). The bottom two panels show the ratio of detection probabilities for the foreground versus both, and the background versus both. Noise models ‘a’, ‘d’ & ‘e’, which we focus on, are highlighted with dashes. The DP for the foreground is effectively always higher than that of the background (or combination). Note that for the ‘back’ and ‘both’ signals, the noise models ‘c’ & ‘f’ nearly perfectly overlap.

the foreground reaches 50% DP in  $\sim 15$  yr.

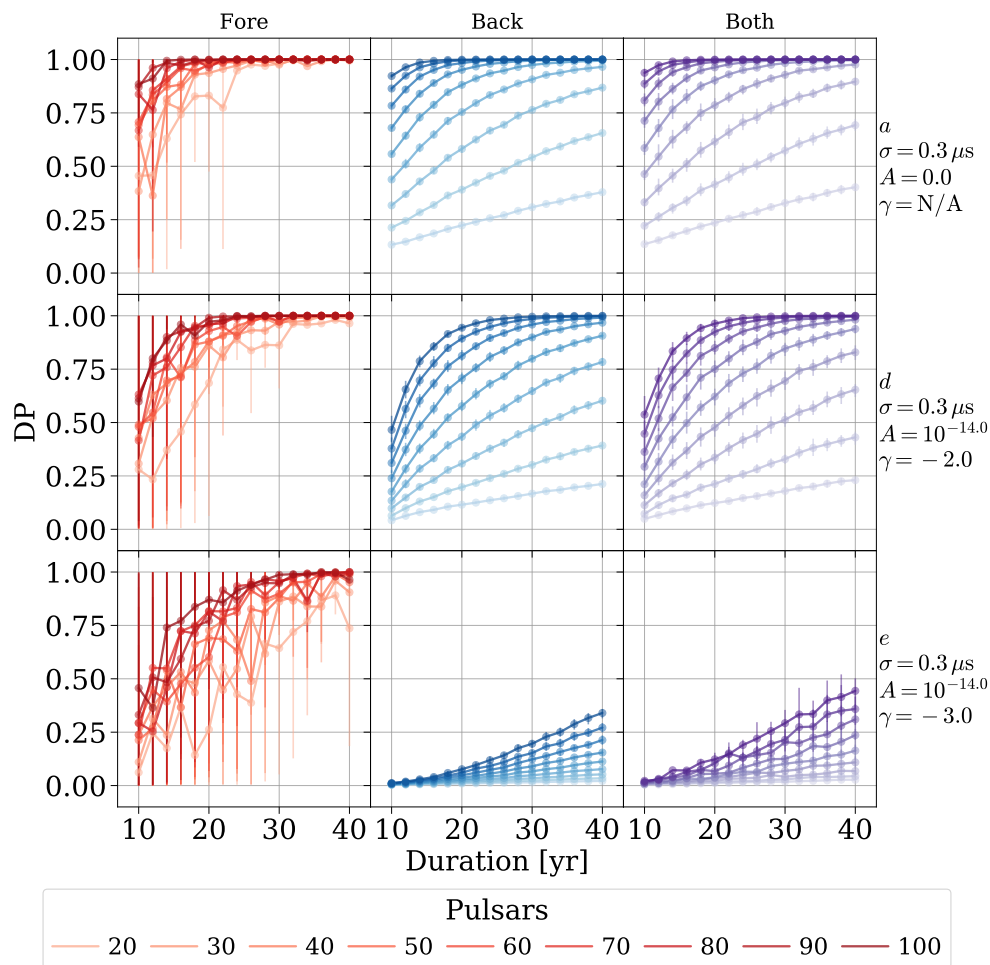
A comparison between the DP curves for varying eccentricity models and varying loss-cone parameters are shown in Figs. 3.16 & 3.17. In general the DP varies by  $\sim 15 - 25\%$  based on varying evolutionary parameters. In the case of the highest eccentricities,  $e_0 = 0.95$  &  $0.99$ , the GWB DP drops drastically, and in the latter case is effectively unobservable for all of the noise and PTA models considered\*. For low to moderate eccentricities ( $e_0 \lesssim 0.75$ ) the time to reach a given DP typically varies by  $\sim 5$  yr, while for varying environmental coupling it varies by up to  $\sim 10$  yr. Still, the dependence on noise model and number of pulsars tends to be more significant than between evolutionary parameters.

The time to reach a given DP is plotted versus number of pulsars in Fig. 3.11. It typically takes  $\sim 5 - 10$  yr to increase from a 50% DP to 90%, but for low numbers of pulsars it can be as long as  $\sim 20$  yr. As already discussed, GWB detection relies on correlations between pulsars which introduces a very strong dependence on the number of pulsars, which is clearly apparent. In the modern ‘d’ noise model, doubling the number of pulsars from 40 to 80 decreases the time to detection by almost a factor of 3: from 65 to 22 years (for DP = 0.9). GWF detection is much less sensitive to number of pulsars, with the time to detection decreasing from  $\sim 23$  to 15 yr from the same increase in pulsar number.

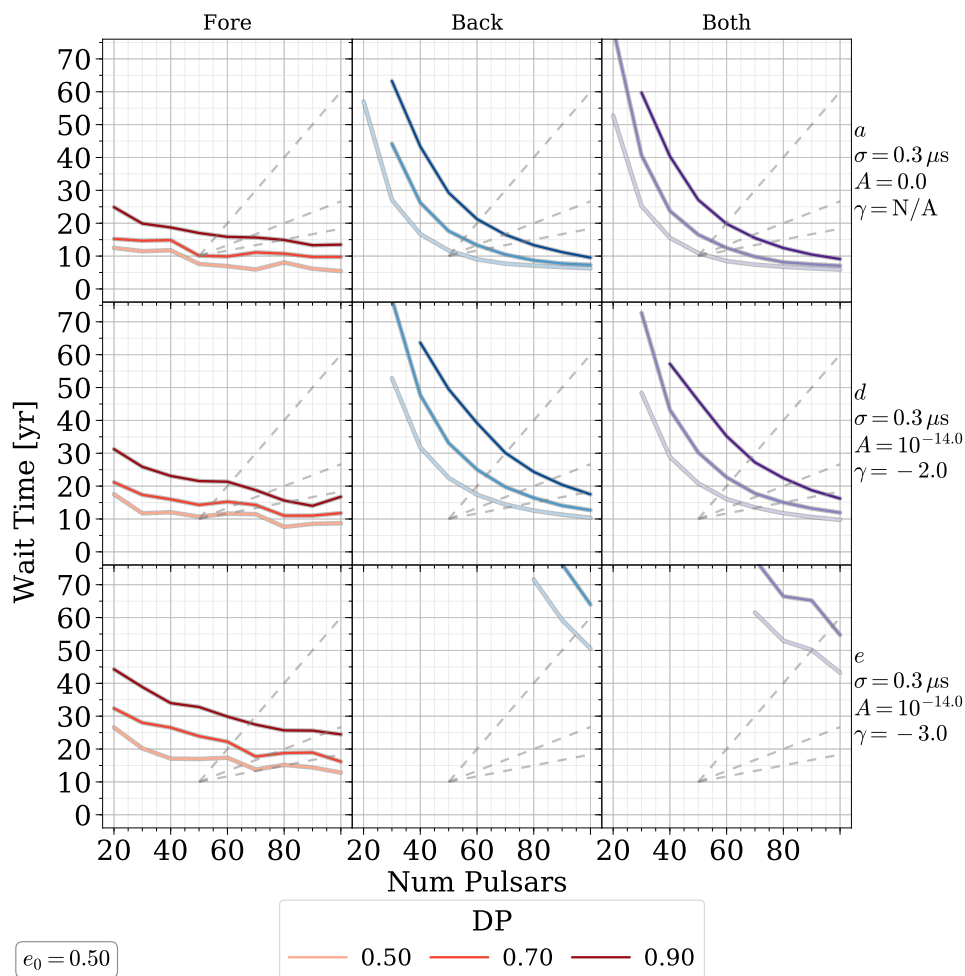
Recall that the IPTA, for example, is expected to expand by roughly 6 pulsars per year, meaning that as it continues to collect data it not only moves upward in this figure, but also to the right. The grey dashed lines assume a starting point of 50 pulsars after 10 years of observations (based on the first IPTA data release) with the addition of 1, 3 & 6 pulsars per year shown in each line. If the 3 pulsars/year line accurately reflects the IPTA expansion, after taking into account the decreased leverage of newly added pulsars, then we would expect the GWF & GWB to reach 90% DP after roughly 18 & 22 yr respectively of total observing time (i.e. 8 & 12 yr of additional observations) for the moderate noise model ‘d’.

---

\*The highest eccentricities decimate the GW spectrum because i) the energy from each binary is spread so broadly, and ii) because the GW hardening rate is drastically increased. This is discussed in detail in [Kelley et al. \(2017\)](#).



**Figure 3.10:** Detection probability versus time for the foreground, background, and combination (‘both’; columns), showing the representative noise-models ‘a’, ‘d’ and ‘e’ (rows). PTA with varying numbers of pulsars are illustrated with differing line colors. The number of pulsars very strongly effects the DP of the GWB (‘back’ and ‘both’) because they rely on cross-correlations to make detections, while the GWF is much less sensitive to pulsar number. The GWF DP is also much more robust against worsening red noise as the foreground spectrum is flatter than that of the GWB.



**Figure 3.11:** Time to reach a given detection probability versus number of PTA pulsars. Optimistic ‘a’ (white noise only), moderate ‘d’, and pessimistic ‘e’ noise models are shown. The grey dashed lines show sample expansion rates for a PTA starting with 50 pulsars after 10 years of observations and adding 1, 3 & 6 pulsars per year. If we consider the 3/yr expansion rate with the ‘d’ noise model realistic for the IPTA, then we would expect to reach 90% DP in roughly 8 & 12 yr (overall observing baselines of 18 and 22 yr) for the GWF & GWB respectively.

## 3.4 DISCUSSION

### 3.4.1 CAVEATS

Some caveats should be borne in mind when examining our results. First, the variations in detection probability between realizations of our foreground models are significant. We have examined the particular binaries which are ‘detected’ in our analysis and find that they are sampled from a fairly broad distribution of parameters. This is reassuring as the statistics are not dominated by a handful of systems. Still, the overall population of MBHB from Illustris may be insufficient to properly sample the full light-cone of observations. Similarly, while we resample our populations to extrapolate to the volume of the universe and include Poisson noise, we are still subject to the same intrinsic systematics of our underlying source population. This is all the more true in a GWF analysis which is by definition far more sensitive to individual sources than the GWB is.

It is also important to note that our mock PTA models are suboptimal because they use uniform pulsar parameters across the array, as required by the detection statistics we use. The time sampling and noise models we have used in our analysis are representative of the published specifications of PTA pulsars. Still, real PTA include highly heterogeneous populations of pulsars with varying sampling and noise characteristics. For example, our pessimistic noise model (‘e’) is consistent with that of some pulsars, but in a heterogeneous PTA, pulsars with such strong noise would contribute far less to the SNR than other pulsars with better noise levels. The pessimistic model ‘e’ is thus certainly *overly* pessimistic. While many pulsars have no observed red noise, our optimistic model ‘a’—with only white noise—is, similarly, likely *overly* optimistic\*. In addition to uncertainties in the most representative noise model, even choosing an accurate number of pulsars is non-trivial. PTA continue to expand by adding new systems, which are a very important part of detection forecasts (e.g. [Taylor et al. 2016b](#)), but have different frequency sensitivities and thus different leverages on SNR.

Finally, the GWF detection statistics we use are suboptimal in at least two respects. First, the statistics in [Rosado, Sesana & Gair \(2015\)](#) do not account for eccentric effects, and we treat the GW energy from single sources that is split across multiple frequency bins as entirely independent. Secondly, the excess power statistics may not perform as well on data where noise processes are harder to distinguish from single GW sources. The presence of line-like noise, due to uncertainties in planetary ephemerides (and their harmonics) for example; or instrumental effects, could introduce significant interference. At the very least, it is likely that foreground sources with periods comparable to the observing duration will be hard to distinguish from red noise when using single pulsar searches. This may necessitate correlated searches with lower recovery efficiencies.

### 3.4.2 CONCLUSIONS

Single GW sources resolvable above the background, which we refer to as **GW Foreground (GWF) sources, tend to be most detectable at frequencies near  $\sim 0.1 - 0.3 \text{ yr}^{-1}$** . At higher frequencies there are fewer sources which also tend to be less loud. At lower frequencies the gravitational wave background (GWB) from all other sources is more likely to drown out the singles. According to most of our models, we would expect there to be a foreground source with a characteristic strain of about  $10^{-15}$  (or timing residual of  $\sim 30 \text{ ns}$ ) after 10 yr of observations. These values are roughly a factor of two higher than those predicted by [Sesana, Vecchio & Volonteri \(2009, e.g. Fig. 3\)](#) and [Ravi et al. \(2014, e.g. Fig. 4\)](#)

\*At least when compared to a heterogeneous PTA with the same total number of pulsars.

When taking eccentricity into account, the primary effect of non-circular evolution is to shift GW energy from lower to higher frequencies. Thus, **increasing eccentricity decreases the occurrence rate of GWF sources**, especially at lower frequencies, both by increasing the GWB amplitude and by diffusing the single source strains themselves. Changes in the effectiveness of stellar scattering, and more generally **the rate of environmental hardening, has a small effect on the properties and occurrence rates of the GWF**. Measurements of the number of GWF sources can thus provide strong constraints on the eccentricity distribution of the underlying MBHB population. Even in the absence of foreground detections, measuring the number of sources contributing to the background could provide the same information. This could be done either by directly resolving numerous loud sources in particular frequency bins (i.e. in frequency space), or indirectly by inference from the degree of anisotropy of the GWB (i.e. in angular space).

**Detection probabilities are usually higher for the GWF than for the GWB** — indicating that the background may not be detected first, as has generally been expected. We emphasize, however, that there is a large variance between realizations in our simulations suggesting that *our population of MBHB may not be large enough for fully converged results*. Based on our models, however, mock PTA comparable to the IPTA are able to reach high detection probabilities in 18 & 22 yr of total observing time for the GWF & GWB respectively, with moderate parameter assumptions\*. Our detection probabilities for the GWB are closely in line with those of [Rosado, Sesana & Gair \(2015\)](#), while our GWF values are notably higher for PTA with  $\sim 50$  pulsars, though again consistent at  $\sim 100$  pulsars. The higher GWF values may be due to our higher single source strains, or possibly a tendency for our sources to reside at lower redshifts owing to our dynamical binary evolution which is not included in the previous, semi-analytic models.

**Pulsar red-noise models have a very strong effect on detection probabilities and times to detection**, especially for the GWB. Comparing between white-noise only, and a moderate red-noise model, the time-to-detection can increase by 50 – 100%. In the case of uniformly severe red noise, prospects for detecting the GWB can become bleak, but many of the currently monitored PTA pulsars show no signs of red noise at all. That being said, **the GWF is much less sensitive to red noise** as single sources are best detected at intermediate frequencies, unlike the GWB which, in our models, is almost always strongest at the lowest accessible frequency bins.

Varying eccentricity and environmental coupling have moderate effects on times to detection. For our intermediate noise model, there is a  $\sim 5$  yr **difference in times to detection between nearly-circular and moderately-high eccentricity**. In the case of single sources, circular systems take less time to be detected, while for the GWB, moderately-high eccentricities ( $e_0 \sim 0.75$ ) are easiest to detect. **Between best and worst case stellar-scattering, there is a difference of  $\sim 10$  yr in times to detection**. The foreground is detectable more quickly as scattering becomes more effective, while the background is optimally detected at moderately-high scattering efficiency ( $\mathcal{F}_{\text{refill}} \sim 0.8$ ).

Red-noise models must be included when constructing realistic forecasts for PTA detection prospects, which has not been done in existing studies. Furthermore, we hope that the red-noise characteristics of PTA pulsars will be more thoroughly explored in the context of the IPTA to better calibrate our expectations. The development of more flexible detection statistics for low-frequency GW single sources would also be very helpful in constructing realistic PTA models and testing them with cosmological MBHB populations.

Many studies have shown that the GWB may be within a decade or so of detection, which is consistent with the results presented here. For the first time, however, we find that the GWF might be just as

\*For comparison, median observing baselines for the IPTA are currently  $\sim 10$  yr.



detectable or possibly even more so. Similarly, PTA upper limits on GWF sources should also be used to constrain the MBHB population as is being done with GWB upper-limits. To this end, additional studies, with larger populations of MBHB, should be explored. Prospects for PTA detection of low-frequency GW seem very promising, and even with only upper limits, we stand at the precipice of making substantial progress in our understanding of MBH binaries and their evolution.

#### ACKNOWLEDGMENTS

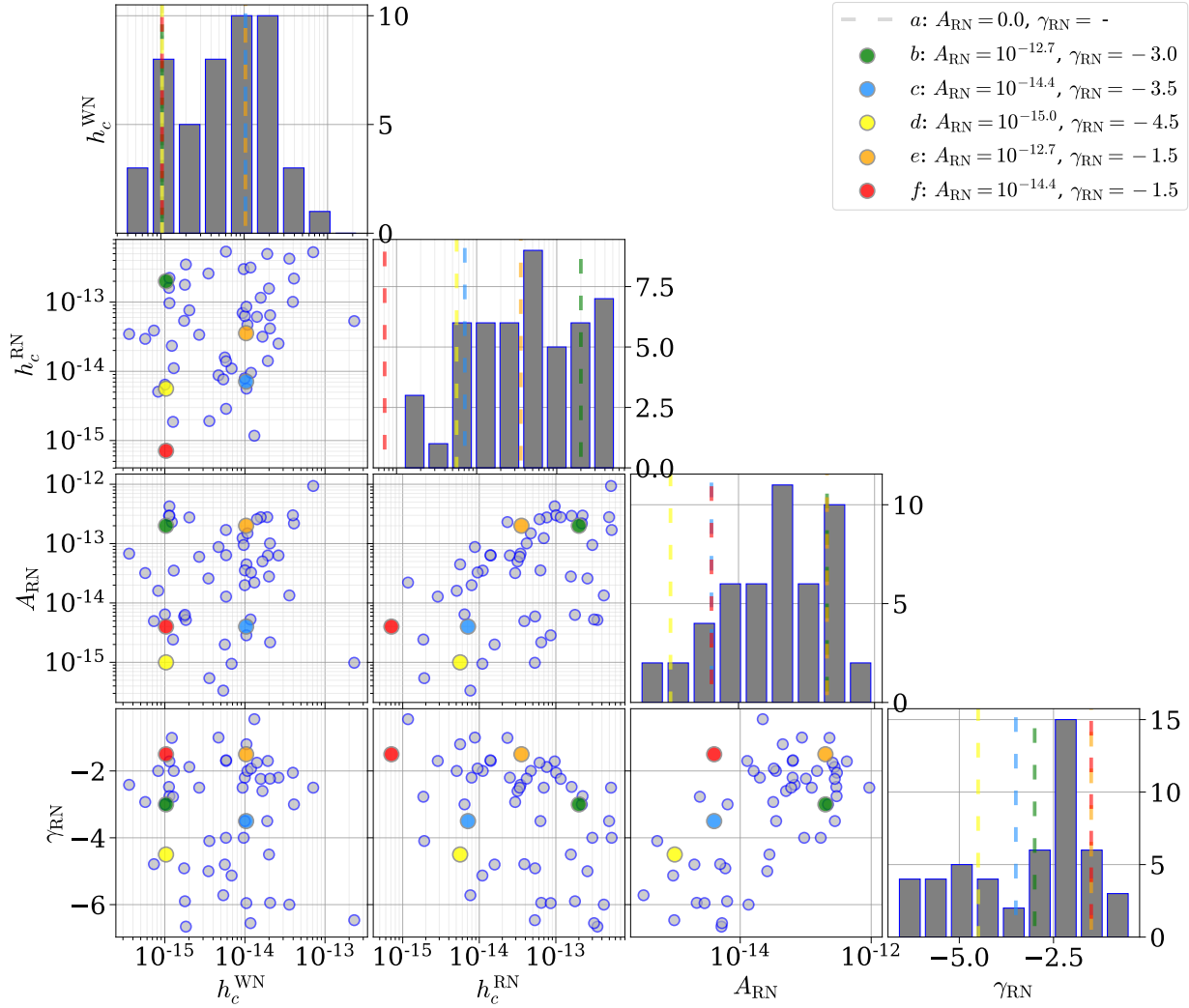
Many of the computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University. We greatly appreciate the help and generous support provided by the Research Computing Group and by the Computational Facility at the Harvard-Smithsonian Center for Astrophysics. This research made use of **Astropy**, a community-developed core Python package for Astronomy ([Astropy Collaboration et al. 2013](#)), in addition to **SciPy** ([Jones et al. 2001](#)), **ipython** ([Prez & Granger 2007](#)), & **NumPy** ([van der Walt, Colbert & Varoquaux 2011](#)). All figures were generated using **matplotlib** ([Hunter 2007](#)).

Luke Zoltan Kelley and Stephen R. Taylor are NANOGrav members. The NANOGrav project receives support from the National Science Foundation (NSF): NSF Physics Frontier Center award number 1430284.

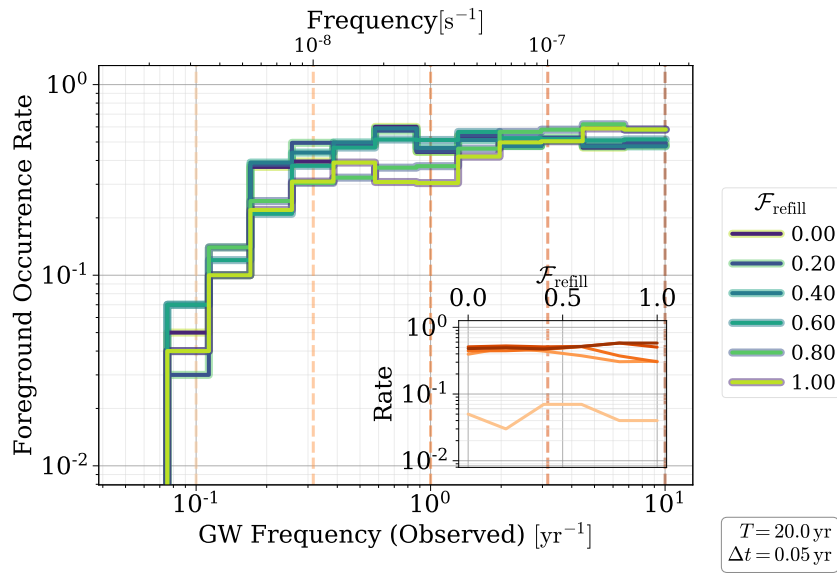
The authors wish to thank the anonymous referee for providing helpful feedback which improved the clarity of this paper.

SUPPLEMENTAL MATERIAL

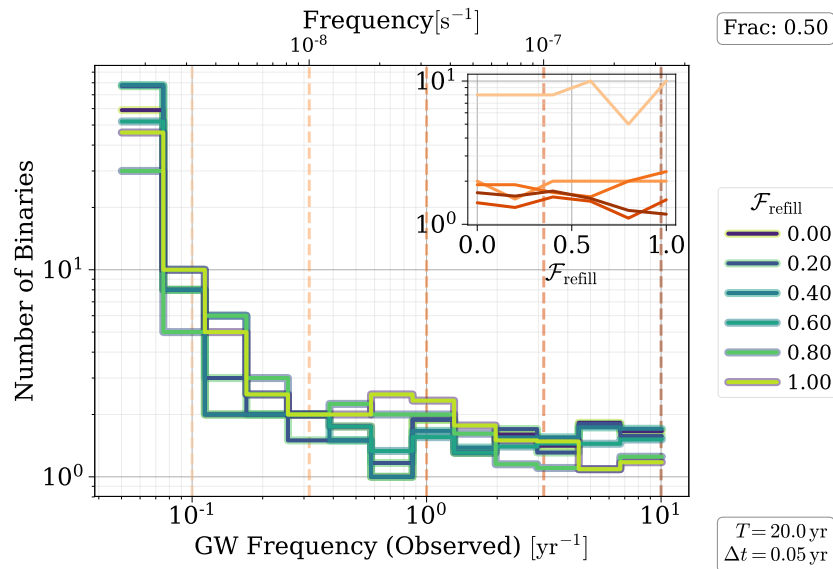
ADDITIONAL FIGURES



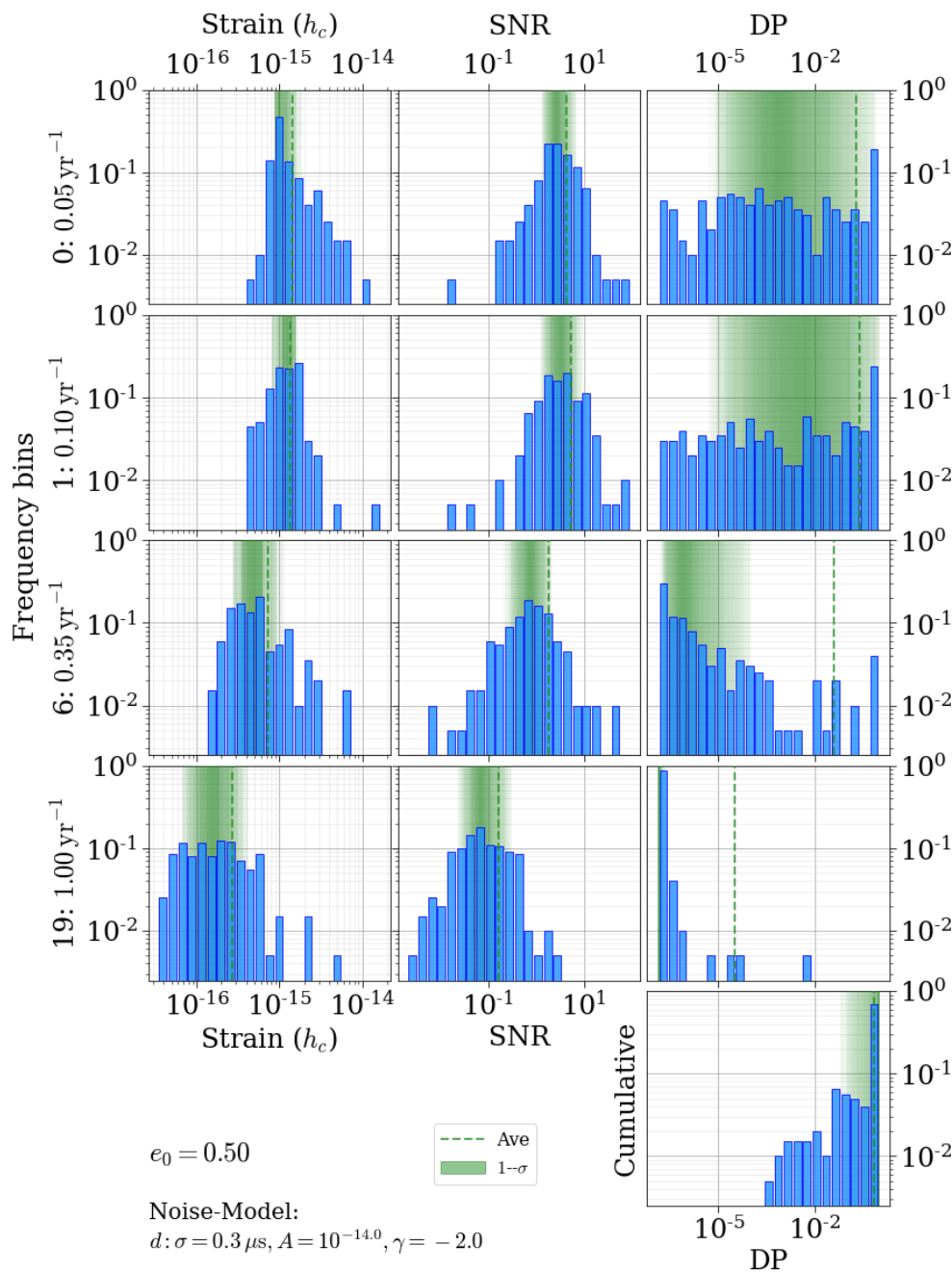
**Figure 3.12:** Noise parameters for all pulsars with measured red-noise characteristics in the PTA public data releases (Desvignes et al. 2016; Verbiest et al. 2016; The NANOGrav Collaboration et al. 2015; Reardon et al. 2016; Caballero et al. 2016; Lentati et al. 2016). When multiple PTA have independent red-noise fits we simply include each characterization independently as they often differ substantially. The first two columns show the white- and red- noise amplitude in units of dimensionless strain at  $f = 1/(10 \text{ yr})^{-1}$ . The third and fourth columns show the red-noise amplitude and spectral index (see: Eq. 3.8).



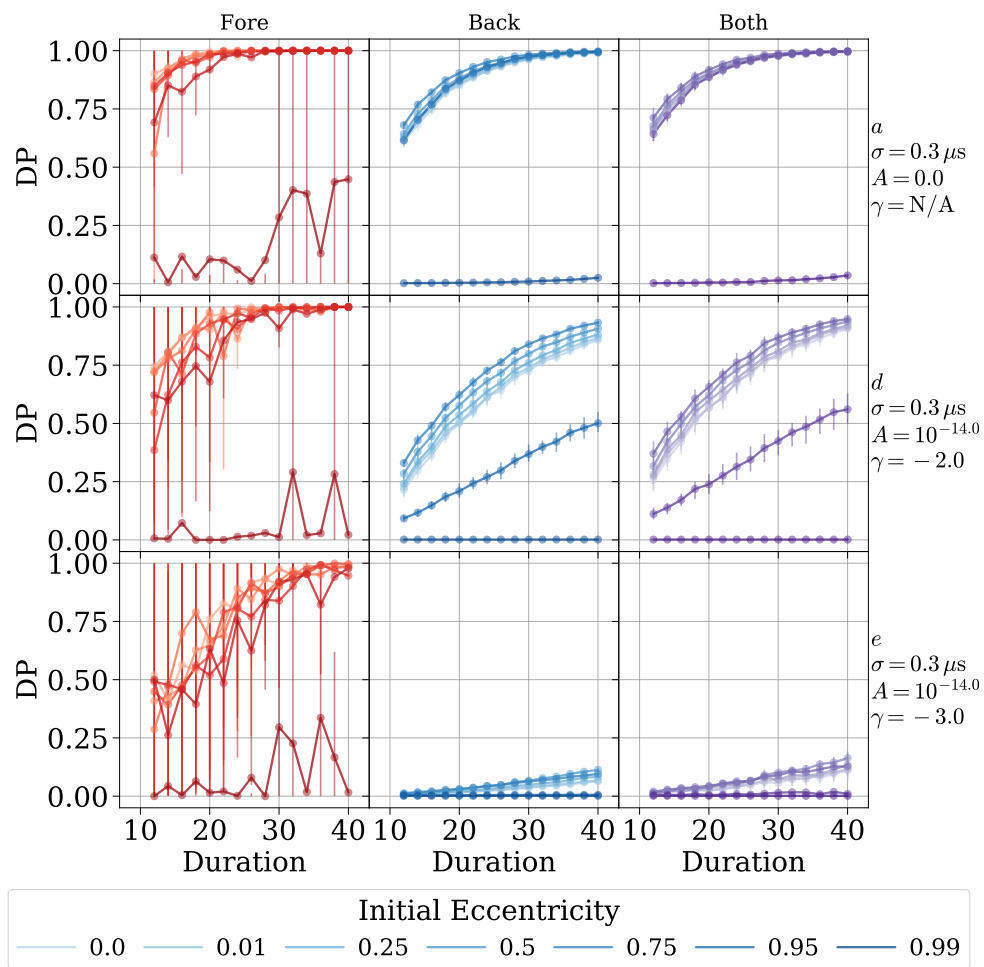
**Figure 3.13:** The occurrence rate of foreground sources per frequency bin for each loss-cone efficiency model, using  $\lambda_{\text{fore}} = 0.5$ . The inset panel shows the foreground occurrence rate versus eccentricity for the highlighted frequencies (dashed orange lines). While the occurrence rate of foreground sources drops significantly at frequencies below  $\sim 1/(5 \text{ yr})^{-1}$ , it remains relatively constant at higher frequencies, unlike in eccentric models. There is also almost no dependence of the foreground rate with stellar scattering efficiency ( $\mathcal{F}_{\text{refill}}$ ).



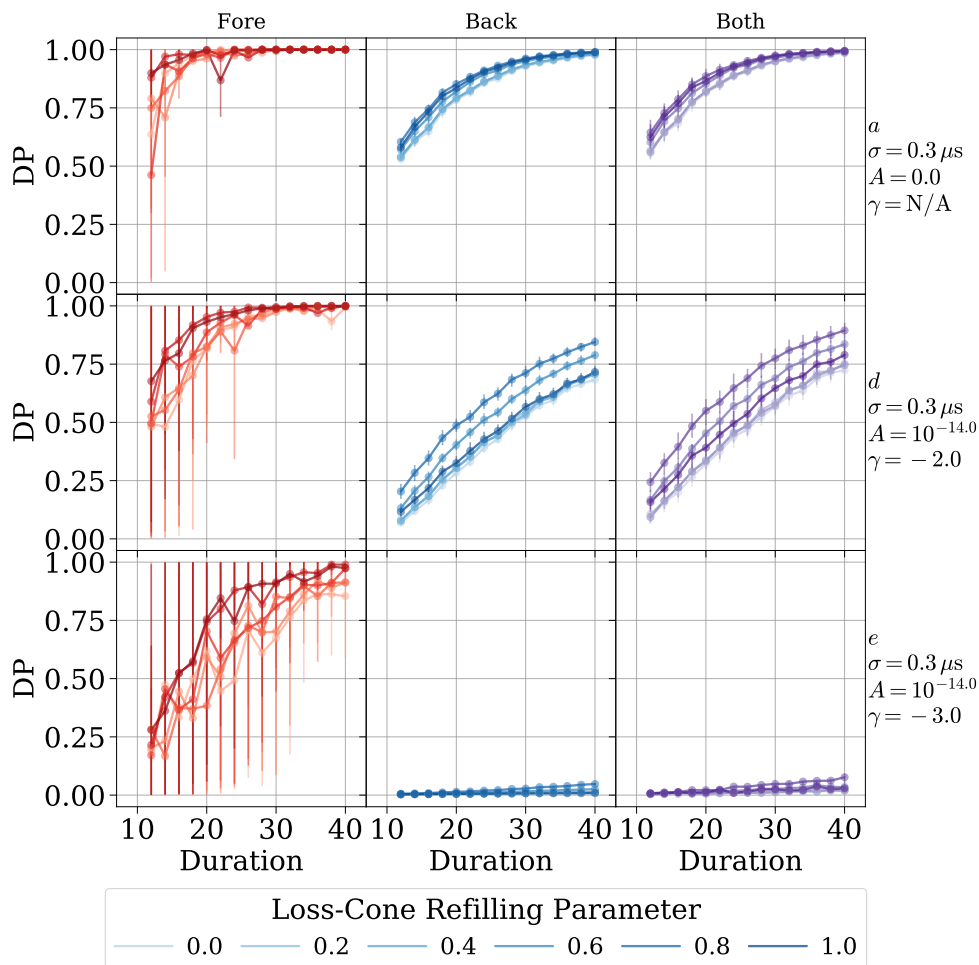
**Figure 3.14:** Number of binaries contributing 50% of the GW energy for each stellar scattering model. The inset panel shows trends versus scattering efficiency at each of the highlighted frequencies (dashed orange). Overall, there is no significant dependence on the relative contribution of binaries from varying stellar scattering.



**Figure 3.15:** Detection statistics for the GWF at different frequencies. The  $e_0 = 0.5$  model is shown, observed by a PTA with 60 pulsars and noise-model ‘d’. Distributions calculated over 200 realizations are shown in blue, with  $1\sigma$  contours in green, and the average values marked by the dashed green line. The strain from each frequency bin, along with the GWB strain and the pulsar noise characteristics, determines the SNR which then maps to a detection probability. The cumulative DP of detecting at least one single-source, over all frequency bins, is shown in the lower right.



**Figure 3.16:** Detection probability is shown versus time for a PTA with 60 pulsars for each eccentricity model. There is mostly a moderate dependence of DP on eccentricity model, causing a variation of  $\sim 5$  yr to reach a given DP. For the most extreme,  $e_0 = 0.99$ , case for the GWF, and additionally the  $e_0 = 0.95$  case for the GWB, the DP drops drastically. Overall, the red-noise model still has a larger effect on DP.



**Figure 3.17:** Detection probability is shown versus time for a PTA with 60 pulsars for each stellar scattering model. There is a moderate dependence of DP on scattering efficiency, causing a variation of  $\sim 5 - 10$  yr to reach a given DP. Overall, the red-noise model still has a larger effect on DP.

# 4

## MASSIVE BH BINARIES AS PERIODICALLY-VARIABLE AGN

*This thesis chapter is in preparation for submission as*  
Luke Zoltan Kelley, Zoltan Haiman, Alberto Sesana, Lars Hernquist

### ABSTRACT

MASSIVE black-hole (MBH) binaries, which are expected to form following the merger of their parent galaxies, produce gravitational waves which should be detectable by Pulsar Timing Arrays at nanohertz frequencies (year periods). While no confirmed MBH binary systems have been seen in electromagnetic observations, a large number of candidates have recently been identified in optical surveys of AGN variability. Using a combination of cosmological, hydrodynamic simulations; comprehensive, semi-analytic binary-merger models; and analytic AGN spectra and variability prescriptions; we calculate, for the first time, the expected electromagnetic detection rates of MBH binaries as periodically variable AGN. Our models predict that many MBH binaries should be present and distinguishable in the existing data. Current binary candidates may have some doppler-boosted variables and likely many hydrodynamically-induced variables. Doppler systems are likely to be seen in only the most massive AGN ( $M_{\text{tot}} \gtrsim 10^9 M_{\odot}$ ), at long orbital periods ( $\tau_{\text{orb}} \gtrsim 3 \text{ yr}$ ); while hydrodynamic systems can extend to lower masses ( $M_{\text{tot}} \gtrsim 10^7 M_{\odot}$ ) and shorter timescales ( $\tau_{\text{orb}} \gtrsim 0.3 \text{ yr}$ ). We present the expected selection biases for observing variable-AGN MBHB, and the expected detection rates for future instruments such as LSST.



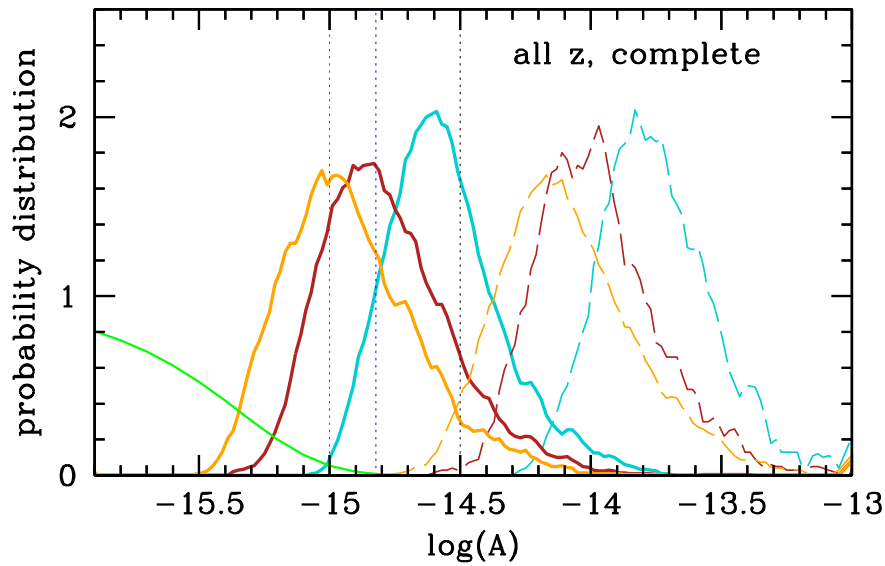
## PREFACE TO CHAPTER 4

A series of papers came out in 2015 looking for signatures of MBH binaries in optical surveys of AGN. [Graham et al. \(2015a\)](#), [Charisi et al. \(2016\)](#), and [Liu et al. \(2016\)](#) searched for periodic variability in AGN light-curves which could indicate the presence of a MBHB in the system. The first two studies put forward populations of ‘candidate’ binary systems—showing periodic signatures, on the order of year periods, in excess of that expected from red-noise, known to pollute AGN light curves—and the third study placed upper limits, consistent with the candidate rates of the others. These candidates, put forward for further study and followup to confirm or reject their binary nature, present an exciting opportunity to identify binaries in the GW emitting regime, for the first time.

While no individual binary candidate would be expected produce GW signals detectable by current Pulsar Timing Arrays (PTA), in [Sesana et al. \(2017b\)](#) we explored the expected, cumulative GW Background (GWB) implied by the detected population. The goal of the study was to determine if the population of candidates as a whole were consistent with PTA upper-limits. Two key aspects of the analysis are 1) taking into account possible selection biases in the inferred MBH masses, and 2) using realistic models of the binary mass-ratio distribution which has no observational constraints at these small separations. Issue (1) is motivated by the observations of [Shen et al. \(2008\)](#) which find that the  $M$ - $\sigma$  relationship, which forms the crux of most MBH mass estimates, may preferentially be calibrated to MBH with masses abnormally large, possibly by as much as  $\approx 0.4$  dex. Given this caveat, the typical mass-ratio distribution of MBH binaries should roughly follow that of galaxy mergers (or at least their bulge velocity dispersion), but there may be strong restrictions on which mass-ratios are able to harden effectively to the small separations corresponding to the observed candidates.

The figure above, from [Sesana et al. \(2017b\)](#), shows the key results of the study. The three solid lines correspond to models which de-bias the MBH mass estimates, and correspond to three different models of mass-ratio ( $q$ ). The blue curve corresponds to a high- $q$  distribution: a log-normal centered around  $q = 1$ ; the red curve to the fiducial model peaked at  $q \approx 0.1$ , with a cutoff at  $q = 0.05$ ; and finally the orange curve corresponds to a *low* -  $q$  distribution, which is the same as the fiducial but without a lower cutoff. The most stringent PTA upper limit, at an amplitude of  $10^{-15}$ , rules out the high- $q$  distribution, and is in strong and moderate tension with the fiducial and low- $q$  distributions respectively. While these mass-ratio distributions are well-motivated, it is worth stating that the high- $q$  distribution is most consistent with the results found throughout this thesis work. Additionally, there are numerous factors to suggest that the GWB amplitudes presented in this work are conservative estimates.

The key result of [Sesana et al. \(2017b\)](#) is that the complete population of candidates is in tension with existing PTA upper limits. The following chapter is motivated by the same population of candidates, but explores the opposite strategy: instead of modeling the expected GWB From the candidates, we use our modeled populations to predict the occurrence rate of periodically-variable candidates. This is the first study, to our knowledge, which makes predictions for rate of electromagnetically observable, close-period MBH binaries.



The distribution of GWB amplitudes implied by the population of periodic-variable binary candidates identified in the CRTS survey (Graham et al. 2015a), taken from Sesana et al. (2017b). The dashed lines correspond to the masses directly inferred for each AGN, while the solid lines account for bias and decrease the mass distributions. The different colored lines correspond to different mass-ratio model: orange, red, teal corresponding to mass-ratio biased low, flat, and biased high. The dashed vertical lines indicate the current, most-stringent PTA upper limits on the GWB amplitude.

## 4.1 INTRODUCTION

Active Galactic Nuclei (AGN) are known to often be triggered by interactions and mergers between their host galaxies (e.g. Comerford et al. 2015; Barrows et al. 2017; Goulding et al. 2018) which drive large amounts of gas towards the galaxy cores and Massive Black-Holes (MBH) within (Barnes & Hernquist 1992). Many examples of “dual-AGN”, pairs of observably accreting MBH in the same system, have been identified in radio, optical and X-ray surveys (e.g. Koss et al. 2012; Comerford et al. 2012). After a galaxy merger, the two MBH are expected to sink towards the center of the post-merger galaxy due to dynamical friction, which is very effective on  $\sim 10^3$  pc scales (Begelman, Blandford & Rees 1980; Antonini & Merritt 2012). Once the MBH reach  $\sim$ pc separations and smaller, and eventually become gravitationally bound as a MBH Binary (MBHB), the continued merging of the system depends sensitively on individual stellar scatterings extracting energy from the binary (e.g. Sesana, Haardt & Madau 2007; Merritt, Mikkola & Szell 2007).

The effectiveness of stellar scattering in ‘hardening’ MBH binaries remains unresolved. Of particular interest is whether and which systems are able to reach the  $\sim 10^{-3}$ – $10^{-1}$  pc separations at which point Gravitational Wave (GW) emission can drive the system to coalesce within a Hubble time (Begelman, Blandford & Rees 1980). While many dual-AGN systems have been observed, there are no confirmed examples of AGN in binaries. If MBHB are able to reach periods of  $\sim$  yr (frequencies  $\sim$  nHz), their GW emission should be detectable by pulsar timing arrays (PTA; Hellings & Downs 1983; Foster & Backer 1990)—the European (EPTA, Desvignes et al. 2016), NANOGrav (The NANOGrav Collaboration et al. 2015), Parkes (PPTA, Reardon et al. 2016), and the International PTA (IPTA, Verbiest et al. 2016). The most recent and comprehensive MBHB merger models suggest that PTA will plausibly make a detection within roughly a decade (e.g. Taylor et al. 2015; Rosado, Sesana & Gair 2015; Kelley et al. 2017), and indeed, the most recent PTA upper-limits on GW signals—particularly on the presence of a power-law, Gravitational-Wave Background (GWB) of unresolved, cosmological sources—have already begun to inform the astrophysical models (Simon & Burke-Spolaor 2016; Taylor, Simon & Sampson 2017).

MBH Binaries form on sub-parsec scales, which, even using VLBI, can only be spatially resolved in a small, local volume of the universe. Spectroscopic, and especially photometric methods, which don’t require binaries to be spatially resolved, have recently put forward large numbers of binary candidates (Erculeous et al. 2012; Tsalmantza et al. 2011; Graham et al. 2015a; Charisi et al. 2016). The focus of this investigation are binaries and candidates identified by periodic variability in photometric surveys of AGN. In particular, Graham et al. (2015a) find 111 candidates in  $\sim 240,000$  AGN using the CRTS survey; Charisi et al. (2016) find 33 in  $\sim 35,000$  AGN using PTF; and Liu et al. (2016) initially identify 3 candidates in 670 AGN using PanSTARRS, however, none are persistent in archival data.

Excitingly, the connection between EM and GW observations of MBHB has already begun to be leveraged using these photometric-variability candidates. While none of the individual candidate systems can be excluded by PTA measurements, Sesana et al. (2017a) demonstrate that the *population* of MBHB that they imply leads to a GWB amplitude in tension with existing PTA upper-limits.

In this paper, we use MBH binary populations based on the Illustris hydrodynamic, cosmological simulations (e.g. Vogelsberger et al. 2014b; Nelson et al. 2015) coupled with comprehensive semi-analytic merger models (Kelley, Blecha & Hernquist 2017; Kelley et al. 2017) to make predictions for the occurrence rates of periodically variable AGN. In §4.2 we summarize the binary population, the AGN spectra we use to illuminate them, and the models of variability we consider. In §4.3 we present our results, focusing on the detectability of MBHB as periodic AGN, both in terms of expected detection rates, and the fraction of binaries observable as a function of binary parameters.

## 4.2 METHODS

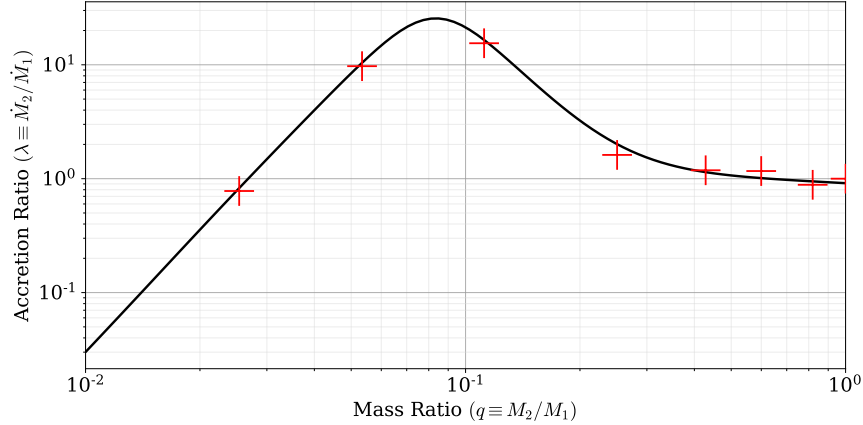
### 4.2.1 MBH BINARY POPULATION AND EVOLUTION

Our MBHB populations are based on the MBH and galaxies in the Illustris simulations. Illustris is an  $(108 \text{ Mpc})^3$  volume of gas-cells and particles representing dark matter, stars, and MBH which is evolved from the early universe to redshift zero (e.g. [Vogelsberger et al. 2014a,b](#); [Genel et al. 2014](#); [Torrey et al. 2014](#); [Rodriguez-Gomez et al. 2015](#); [Nelson et al. 2015](#)). The simulations include sub-grid models for star formation, stellar nucleosynthesis & metal enrichment, and stellar & AGN feedback. MBH particles are initialized with a seed mass of  $\sim 10^5 M_\odot$  in massive halo centers and then grow via accretion of local gas using a Bondi prescription. Details of BH prescription and resulting MBH and AGN populations are presented in [Sijacki et al. \(2015\)](#). In the Illustris simulations, after or during a galaxy merger, once MBH come within  $\sim 10^2 - 10^3$  pc of one-another—roughly their gravitational smoothing length—they are manually merged and moved to the potential minimum of the halo. To more carefully examine the MBHB merger process and dynamics, we ‘post-process’ the MBH mergers using semi-analytic models.

The details of the merger models, and the resulting merger dynamics, are described in [Kelley, Blecha & Hernquist \(2017\)](#); [Kelley et al. \(2017\)](#); here we outline some of the key aspects. MBH-MBH “merger-events” are identified in Illustris on  $\sim$  kpc scales. We then consider each of these events independently by extracting the MBH masses, and spherically-averaged galaxy density and velocity profiles for each mass constituent (DM, stars, gas) of the host. These profiles are then used to calculate hardening rates of the semi-major axis based on prescriptions for dynamical friction ([Chandrasekhar 1942](#); [Binney & Tremaine 1987](#)), stellar ‘loss-cone’ scattering ([Magorrian & Tremaine 1999](#)), viscous drag from a circumbinary disk ([Haiman, Kocsis & Menou 2009](#); [Tang, MacFadyen & Haiman 2017](#)), and GW emission ([Peters & Mathews 1963](#); [Peters 1964](#)). Dynamical friction is required to harden the system on  $10 - 10^3$  pc scales, after which stellar scattering is dominant until the GW-dominated regime on  $\sim 10^{-2} - 10^{-4}$  pc. In some systems, viscous drag is dominant near  $\sim 10^{-2}$  pc.

Both stellar scattering and viscous drag remain highly uncertain processes. The largest uncertainty affecting merger outcomes is likely the effectiveness of stellar scattering: in particular, how efficiently the stellar ‘loss-cone’—those stars able to interact with the binary—are repopulated. Typical coalesce lifetimes are Gyr. Binaries which are both very massive  $M \equiv M_1 + M_2 \gtrsim 10^9 M_\odot$ , and near equal mass ratio  $q \equiv M_2/M_1 \gtrsim 0.1$ , are generally able to coalesce within a Hubble time. Systems with both lower total masses ( $M \lesssim 10^8 M_\odot$ ), and more extreme mass ratios ( $q \lesssim 10^{-2}$ ) usually stall at either kpc or pc separations. The fate of the remaining systems depends more sensitively on the assumed dynamical parameters (i.e. the loss-cone refilling rate). Note that this differs from numerous previous studies finding that more-massive systems merge *less* effectively (e.g. [Yu 2002](#); [Cuadra et al. 2009b](#)).

Predictions for the GWB and its prospects for detection by PTA are presented in [Kelley et al. \(2017\)](#), along with a description of our formalism for eccentric binary evolution. Most models predict GWB amplitudes at periods of 1 yr,  $A_{\text{yr}^{-1}} \approx 0.5 - 0.7 \times 10^{-15}$ , roughly a factor of 2 below current sensitivities, and detectable within about a decade. Predictions for GW signals from individually resolvable ‘single-sources’ are presented in [Kelley et al. \(2018a\)](#), and are comparable in detectability to the GWB. The results we present in this paper are relatively insensitive to variations in binary evolution parameters, compared to those of the electromagnetic and observational models we describe below. For reference, the evolutionary model used here assumes an always full loss-cone and initial binary eccentricities of  $e_0 = 0.5$ .



**Figure 4.1:** Accretion ratio data points are from hydrodynamic simulations of MBHB in circumbinary accretion disks by [Farris et al. \(2014\)](#). The line is a manual fit with the function and parameters in Eq. 4.1.

#### 4.2.2 MBH ACCRETION AND AGN SPECTRA

Our merger models follow the constituent MBH of a given binary for long after it has “merged” in Illustris. We use the accretion rate recorded in Illustris of the single, remnant MBH as a measure of the accretion rate to the binary system as a whole, feeding the circumbinary disk,  $\dot{M} = \dot{M}_1 + \dot{M}_2$ . This leaves an ambiguity in the feeding rate to each individual BH. To resolve this, we use the results from the detailed circumbinary-disk simulations in [Farris et al. \(2014\)](#), which give the ratio of accretion rates for a variety of binary mass-ratios:  $\lambda = \lambda(q) \equiv \dot{M}_2 / \dot{M}_1$ . We have manually fit the simulation data-points with the curve,

$$\lambda = q^{a_1} e^{-a_2/q} + \frac{a_3}{(a_4 q)^{a_5} + (a_4 q)^{-a_5}} \quad (4.1)$$

$$a_1 = -0.25, \quad a_2 = 0.1, \quad a_3 = 50.0, \quad a_4 = 12.0, \quad a_5 = 3.5.$$

We assume that the system is Eddington limited on large scales, i.e.  $\dot{M} \leq \dot{M}_{\text{Edd}} \equiv \dot{M}_{\text{Edd},1} + \dot{M}_{\text{Edd},2}$ , where  $\dot{M}_{\text{Edd}} = 1.4 \times 10^{18} \text{ g s}^{-1} \left(\frac{M}{M_{\odot}}\right) \left(\frac{\epsilon_{\text{rad}}}{0.1}\right)^{-1}$ , and  $\epsilon_{\text{rad}}$  is the radiative efficiency. We let each MBH individually exceed Eddington\* (e.g. [Jiang, Stone & Davis 2014](#)) which can occur for the secondary when  $\lambda > 1.0$ , corresponding to  $q \gtrsim 0.03$ . The secondary accretion rate is maximized at  $\lambda_{\text{max}} = \lambda(q \approx 0.08) \approx 25$ .

The parameters for MBH evolution in Illustris are calibrated to match the observed M- $\sigma$  relation, and the AGN (bolometric) luminosity function based on a constant radiative efficiency of 0.05. (see, [Sijacki et al. 2015](#)). For our analysis, we calculate full spectra for each MBH based on its mass and Eddington ratio  $f_{\text{Edd}} \equiv \dot{M} / \dot{M}_{\text{Edd}}$ . For  $f_{\text{Edd}} \geq 10^{-2}$ , we assume the accretion flow is radiatively efficient and use a [Shakura & Sunyaev \(1973\)](#) ‘thin’-disk solution, which assumes emission is purely thermal from each annulus of the disk. For  $f_{\text{Edd}} < 10^{-2}$  we assume radiatively inefficient accretion in the form of an ADAF ([Narayan & Yi 1995](#)), and use the emission model from [Mahadevan \(1997\)](#). The ADAF model includes self-absorbed synchrotron emission, bremsstrahlung, and inverse-Compton of synchrotron photons. We thus calculate AGN spectra as,

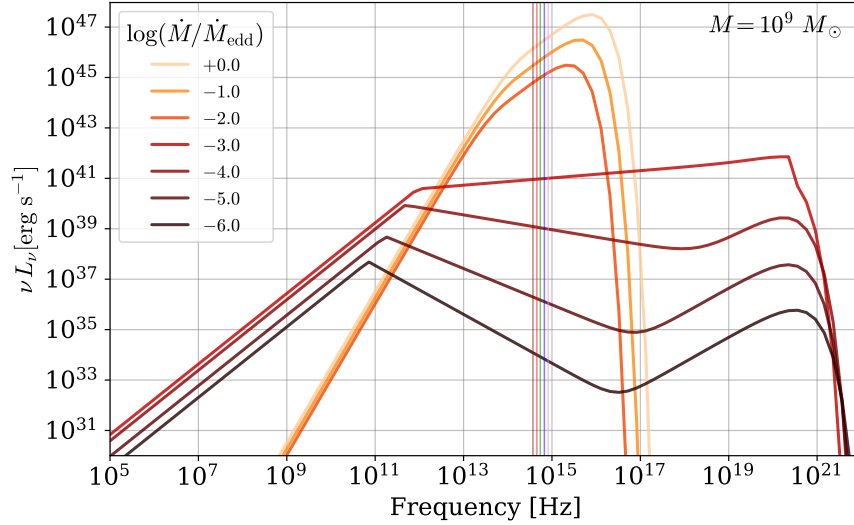
$$F_{\nu} = F_{\nu}^{\text{thin}}(M, f_{\text{Edd}}) \quad f_{\text{Edd}} \geq 10^{-2}, \quad (4.2)$$

$$= F_{\nu}^{\text{ADAF}}(M, f_{\text{Edd}}) \quad f_{\text{Edd}} < 10^{-2}. \quad (4.3)$$

Spectra for a variety of accretion rates onto a  $M = 10^9 M_{\odot}$  BH are shown in Fig. 4.2. Bolometric and b-band luminosities<sup>†</sup> are shown in Fig. 4.3, along with the effective radiative-efficiency ( $L_{\text{bol}}/L_{\text{Edd}}$ ) and

\*We also explore the effects of Eddington limited accretion for both MBH in §4.4.2.

<sup>†</sup>We show the b-band ( $\lambda = 445\text{nm}$ ) for comparison with [Hopkins, Richards & Hernquist \(2007\)](#), while our anal-



**Figure 4.2:** AGN spectra for an MBH of mass  $M = 10^9 M_\odot$ , and a variety of accretion rates. For Eddington ratios  $f_{\text{Edd}} \equiv \dot{M}/\dot{M}_{\text{Edd}} < 10^{-2}$ , we use the ADAF emission model from [Mahadevan \(1997\)](#), while for larger accretion rates we use a thermal, Shakura-Sunyaev spectrum.

b-band luminosity fraction ( $\nu_b L_b / L_{\text{bol}}$ ).

The Illustris AGN luminosity function, comparing a constant radiative efficiency of 0.05 and integrating our model spectra, in the left-panel of Fig. 4.4. Observationally-based luminosity functions from [Hopkins, Richards & Hernquist \(2007\)](#) are included for comparison. The right panel shows values for spectral luminosity in the b-band. Comparing observed and simulated AGN observations is non-trivial. On the observational side there are extinction, bolometric corrections, K-corrections, and selection biases; while on the simulation side, we are using disk-integrated quantities based on semi-analytic models instead of either radiative transfer calculations or full disk-simulations. None the less, our models agree with observations to well within an order of magnitude.

The luminosity functions from [Hopkins, Richards & Hernquist \(2007\)](#) shown in Fig. 4.4 correct for obscuration. For our analysis, we use the same model which assumes that only a luminosity-dependent fraction of systems are observable,

$$f(L) = f_{46} \left( \frac{L}{10^{46} \text{ erg s}^{-1}} \right)^\beta, \quad (4.4)$$

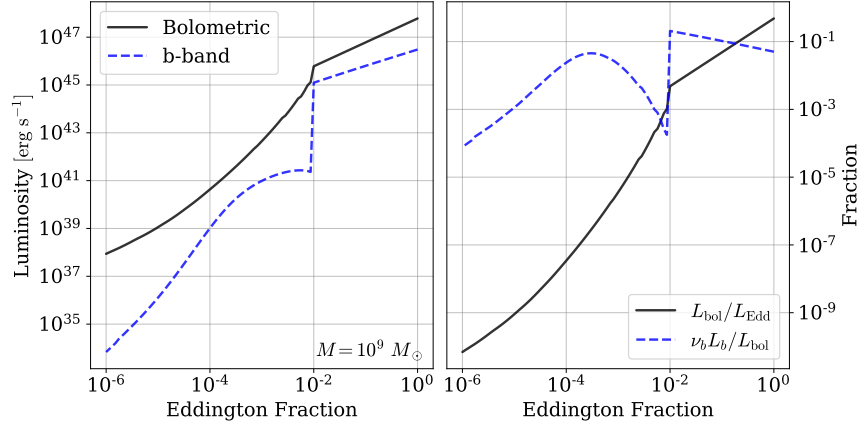
where  $L$  is the bolometric luminosity, and we use the b-band values of:  $f_{46} = 0.260$  and  $\beta = 0.082$ . We convert between bolometric and spectral luminosity using the bolometric correction ([Hopkins, Richards & Hernquist 2007](#)),

$$\frac{L}{L_i} = c_1 \left( \frac{L}{L_\odot} \right)^{k_1} + c_2 \left( \frac{L}{L_\odot} \right)^{k_2}, \quad (4.5)$$

with  $c_1 = 7.40$ ,  $c_2 = 10.66$ ,  $k_1 = -0.37$ , and  $k_2 = -0.014$ .

One additional adjustment is required for spectra from disks in binary systems: the presence of a companion leads to truncation of each disk at a radius comparable to the Hill-radius. Specifically we set

analysis uses the v-band ( $\lambda = 551\text{nm}$ ) for consistency with binary candidate observations (e.g. in CRTS). The change in luminosity between these bands is very minor.



**Figure 4.3:** Luminosity and radiative efficiency versus Eddington ratio. The left panel shows the bolometric and b-band luminosity, calculated for a  $M = 10^9 M_{\odot}$  MBH, against a variety of accretion rates. The right panel gives the overall radiative efficiency  $\epsilon_{\text{rad}} = L_{\text{bol}}/L_{\text{Edd}}$ , as well as the fraction of energy emitted in the b-band (i.e. the inverse of the bolometric correction).

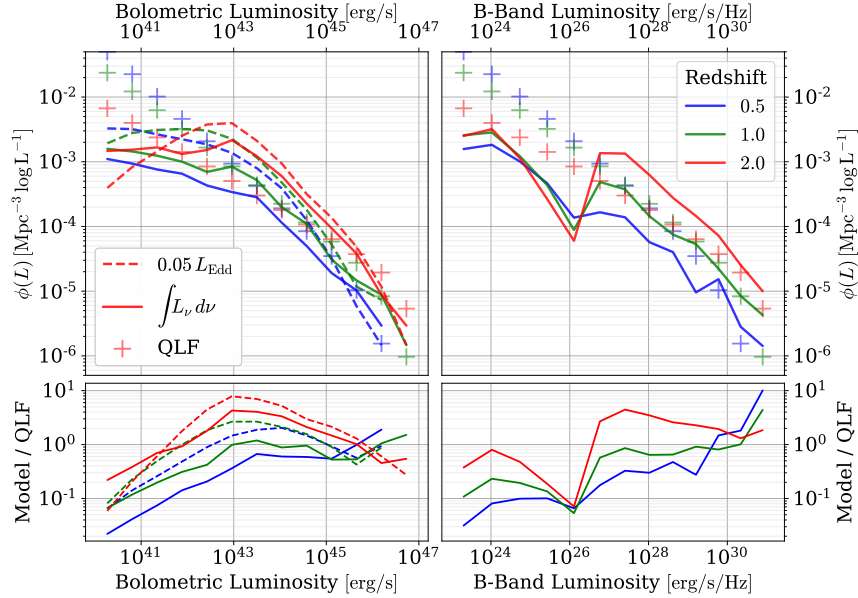
the outer-edge of each disk to,

$$r_{\text{max}} = \frac{a}{2} \left( \frac{M_i}{M} \right)^{1/3}, \quad (4.6)$$

$$\sim 74 R_s \left( \frac{\tau_{\text{orb}}}{5 \text{ yr}} \right)^{2/3} \left( \frac{M}{10^9 M_{\odot}} \right)^{-2/3} \quad (4.7)$$

where  $M_i$  is the mass of the primary or secondary,  $a$  is the semi-major axis,  $\tau_{\text{orb}}$  is the orbital period and  $R_s$  is the Schwarzschild radius of the MBH in question. While the bright, optical emission in AGN tends to come from relatively small radii, disk truncation can be important for especially massive BH and those in short-period binaries\*. For example, the optical luminosity of a  $10^9 M_{\odot}$  MBH, in a 5 yr period binary, can be decreased by  $\sim 10 - 20\%$ .

\*In §4.4.2 we show the effects on truncation of prospective detection rates.



**Figure 4.4:** Comparison of predicted versus observed luminosity functions at three different redshifts,  $z = 0.5, 1.0$  &  $2.0$ . The crosses give values from the observationally-fit quasar luminosity function (QLF) from [Hopkins, Richards & Hernquist \(2007\)](#), while the lines show predictions from our models. In the left panel, bolometric luminosity-functions are compared, while the right panel shows functions in the b-band. The lower panels show the ratio of model to QLF fitting-functions. The solid lines are based on the full spectral models discussed here, while the dashed lines (left panel) use the approximation used in [Sijacki et al. \(2015\)](#) (similar to that used for calibrating Illustris), that the bolometric luminosity is a fixed 5% of the Eddington luminosity.

#### 4.2.3 OBSERVATION PARAMETERS AND CALCULATIONS

We present our results in terms of observability with CRTS- and LSST- like observatories. We adopt parameters based on CRTS in particular, because of their large sample of binary candidates, but we consider this representative of current survey capabilities in general and thus applicable to PTF and PanSTARRS, for example\*. For spectral-flux, we adopt v-band sensitivities of  $F_{\nu, \text{sens}}^{\text{CRTS}} = 4 \times 10^{-28}$  erg/s/Hz/cm<sup>2</sup> and  $F_{\nu, \text{sens}}^{\text{LSST}} = 3 \times 10^{-30}$  erg/s/Hz/cm<sup>2</sup> for CRTS and LSST respectively†. The amount of variability which is detectable in a source is dependent on its signal-to-noise ratio  $\Theta$ , with a floor set by the photometric stability/precision of a given instrument. We parametrize the variability sensitivity as,

$$\delta F_{\text{sens}} = \Theta^{-1} + \delta F_{\text{floor}}, \quad (4.8)$$

where we assume the noise is a factor of five below the sensitivity threshold, i.e.,  $\Theta \equiv 5F_{\nu} / F_{\nu, \text{sens}}$ , and we use  $\delta F_{\text{floor}} = 0.05$  as a fiducial value, based on the smallest variability amplitudes seen in [Graham et al. \(2015a\)](#) ‡.

The overall number-density (i.e.  $\text{Mpc}^{-3}$ ) of sources is the parameter most directly extracted from our

\*From [Charisi et al. \(2016\)](#) and [Liu et al. \(2016\)](#), the sensitivities of PTF and PanSTARRS are  $\sim 2$  and  $\sim 10$  times deeper than CRTS, but both have shorter time baselines in their analyses.

†The CRTS value we get from the cutoff in the flux distribution of candidates from [Graham et al. \(2015a\)](#), while the LSST value is from [Ivezic et al. \(2008\)](#). Our detection rates are not strongly dependent on the particular flux-threshold, as discussed in §4.4.2.

‡The form of (4.8) roughly matches that found with HST by [Sarajedini, Gilliland & Kasm \(2003\)](#), see their Fig. 3) and PanSTARRS by [Liu et al. \(2016\)](#), see their Fig. 4). In both cases, the authors find minimum detectable variabilities of  $\sim 1 - 2\%$ , and then select systems using a cut which is some factor larger, at  $\sim 5\%$ .



models. To better compare with observations, and to reduce the impact of systematic uncertainties in our luminosity functions, we focus on the number of observable binaries as a fraction of the expected number of all observable AGN. For the latter, we calculate the observability of all Illustris MBH, using the same spectral models described in §4.2.1, out to a redshift  $z_{\max} = 2.0$ . For the aforementioned sensitivities, we expect  $4 \times 10^5$  AGN ( $9 \times 10^{-7}$  arcsec $^{-2}$ ) to be observable by CRTS, and  $10^7$  AGN ( $3 \times 10^{-5}$  arcsec $^{-2}$ ) by LSST, assuming a survey area of 33,000 deg $^2$ . For comparison, CRTS actually observes  $\approx 330,000$  spectroscopically-confirmed AGN [Graham et al. \(2015a\)](#), which only differs from our predicted value by 20%. We emphasize, however, that this degree of consistency is largely fortuitous and we would naively assume systematic uncertainties of at least a factor of a few (i.e. Fig. 4.4).

Our entire population of MBHB is  $\sim 10^4$  systems, over all simulation redshifts, while those that reach  $\tau_{\text{orb}} \sim 1$  yr before redshift zero is  $\approx 2,500$ . The redshift distribution of sources at  $\tau_{\text{orb}} = 1$  yr is shown in Fig. 4.10. In numerous of our calculations, the number of observable systems is of order unity, leading to issues of small number statistics. With this number of total binaries, we cannot consider differential observability as a function of redshift. Instead, to each binary we associate a number density which is the inverse of the Illustris volume,  $V_{\text{ill}} = (108 \text{ Mpc})^3$ . We then convert to an effective number density of sources by finding the fraction of comoving volume in the universe (out to  $z_{\max}$ ) in which that binary could be observed,  $f$ . The total number-density of observed sources is then,

$$\phi = \frac{1}{V_{\text{ill}}} \sum_i f_{\text{obs},i}. \quad (4.9)$$

Because we are treating  $\phi$  as a redshift-averaged quantity, one can easily convert between number- and surface- density using the conversion factor 1.2 Mpc $^3$ /arcsec $^2$  (for  $z_{\max} = 2.0$ ).

#### 4.2.4 MODELS OF VARIABILITY

The luminosity of an object in a binary system will not necessarily vary on the orbital period or at all. The premise of photometric identification of MBH binaries is that the binary period is somehow imprinted into variations of the observed luminosity. In the particularly convincing example of PG 1302-102 ([Graham et al. 2015b](#)), sinusoidal variations in the light-curve can be well explained by doppler-boosting from a mildly relativistic orbital velocity ([D’Orazio, Haiman & Schiminovich 2015](#); but see also [Liu, Gezari & Miller 2018](#)). Additionally, purely hydrodynamic modulations to accretion rates have been observed in simulations (e.g. [Farris et al. 2014](#)). Here we describe models for both types of variability mechanisms.

##### 4.2.4.1 DOPPLER

Any component of the binary orbital-velocity along the observer’s line-of-sight will lead to both doppler beaming and a doppler shift in the observed spectrum of the source. In terms of the Lorentz factor,  $\gamma \equiv (1 - v^2/c^2)^{-1/2}$ , and Doppler factor,  $D \equiv [\gamma (1 - v_{\parallel}/c)]^{-1}$ , the observed flux is ([D’Orazio, Haiman & Schiminovich 2015](#)),

$$F_{\nu} = D^3 F'_{\nu'}, \quad (4.10)$$

where the observed frequency  $\nu$  is related to the rest-frame frequency as  $\nu = D \nu'$ . Assuming a power-law of index  $\alpha_{\nu}$  for the section of the spectrum being observed, the doppler variation in flux from the source will be ([Charisi et al. 2018](#)),

$$\frac{\Delta F_{\nu}^d}{F_{\nu}} = (3 - \alpha_{\nu}) \frac{v}{c} \sin i, \quad (4.11)$$

where  $v$  is the orbital velocity,  $c$  is the speed of light, and  $i$  is the angle between the line-of-sight and the orbital angular-momentum vector. The sensitivity of doppler boosting to frequency and thus spectral

shape offers a powerful method of testing it as a variability mechanism. D’Orazio, Haiman & Schiminovich (2015) have shown that, in both the optical and ultraviolet, this model explains the periodic variations observed in PG 1302-102.

In full generality, an AGN spectra may not be a power-law at the frequency of interest, so we construct a full spectrum for each MBH in our simulations and numerically calculate the change in flux using Eq. 4.10. Additionally, the doppler-boosting of each MBH in a binary is necessarily  $\pi$  out of phase, thus we determine the overall system variation as,

$$\delta F_{\nu}^d \equiv \frac{\Delta F_{\nu,1}^d - \Delta F_{\nu,2}^d}{F_{\nu,1} + F_{\nu,2}}. \quad (4.12)$$

For inclination angle, we assume a uniform value of  $\sin i = 0.64$ —to account for selection effects which will prefer doppler-variable systems to be nearer edge-on.

#### 4.2.4.2 HYDRODYNAMIC

Periodic variations in accretion rates are frequently observed in hydrodynamic simulations of circumbinary disks (e.g. Roedig et al. 2012; Farris et al. 2014; Muñoz & Lai 2016; D’Orazio et al. 2016). While significant uncertainties remain in understanding these accretion flows, the general pattern emerging is that three distinct mass-ratio regimes exist. For extreme mass-ratios ( $q \lesssim 10^{-2}$ ), the secondary is a minor perturbation to the circumbinary disk, and the accretion flow remains steady. At intermediate mass ratios ( $10^{-2} \lesssim q \lesssim q_{\text{crit}}$ , where  $q_{\text{crit}} \approx 1/3$ ), a gap is opened by the secondary and the accretion rate onto it varies by a factor of a few, on the binary orbital period.

For near-equal mass ratio systems ( $q \gtrsim q_{\text{crit}}$ ), a highly distorted cavity is evacuated around the binary, out to roughly twice the binary separation. At the outer edge of the cavity, a significant over density of material develops. The keplerian orbital period of that over density sets the variation timescale as 5 – 6 times the binary period. The binaries we are considering (i.e.  $M > 10^6 M_{\odot}$ ,  $\tau_{\text{orb}} \sim \text{yr}$ ) are almost always in the GW-dominated regime in which the hardening timescale—the duration a given binary spends at that separation—decreases rapidly with decreasing orbital period. Thus, if a given variational timescale is probing binaries at shorter periods, the number of observable systems decreases.

In our fiducial models, we assume that all binaries with mass ratios above  $q_{\text{crit}} = 1/3$  are observable at  $\tau_{\text{var}} = 5 \tau_{\text{orb}}$ . The accretion-rate variations in simulations predominantly effect the secondary MBH (e.g. Farris et al. 2014), so we model the overall hydrodynamic variations as,

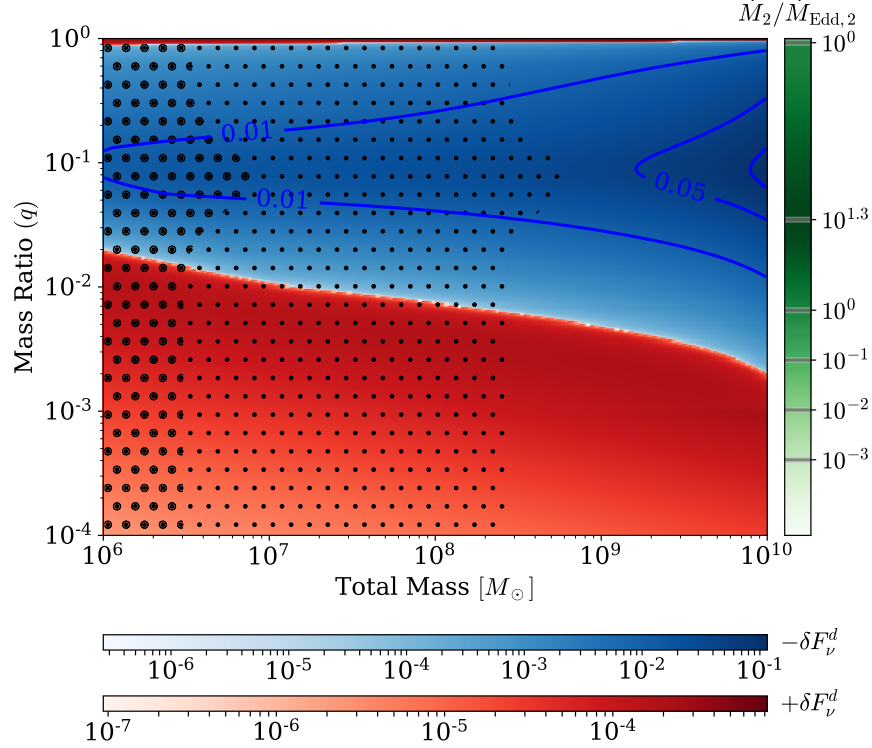
$$\delta F_{\nu}^h \equiv \frac{\Delta F_{\nu,2}^h}{F_{\nu,1} + F_{\nu,2}} = \frac{F_{\nu,2}(M_2, \chi f_{\text{Edd},2}) - F_{\nu,2}(M_2, f_{\text{Edd},2})}{F_{\nu,1} + F_{\nu,2}}, \quad (4.13)$$

where we take the effective enhancement to the accretion rate as  $\chi = 1.5$ . In §4.4.2 we explore alternative values of  $\chi$ , and the importance of the  $\tau_{\text{var}}$  assumption.

## 4.3 RESULTS

### 4.3.1 DOPPLER VARIABLE BINARIES

The doppler-induced variability amplitudes for a grid of mass ratio and total mass, at a period  $\tau_{\text{orb}} = 1 \text{ yr}$ , are shown in Fig. 4.5. The system accretion-rate is set to  $f_{\text{Edd},\text{sys}} = 1$ , and the accretion partition-function,  $\lambda(q)$ , gives the relative accretion rate onto the primary and secondary. The secondary accretion rate is shown in the green color-bar to the right. Following the definition of the doppler variability-amplitude (Eq. 4.12), the secondary’s variations are dominant when negative (blue) and the primary’s when positive (red).



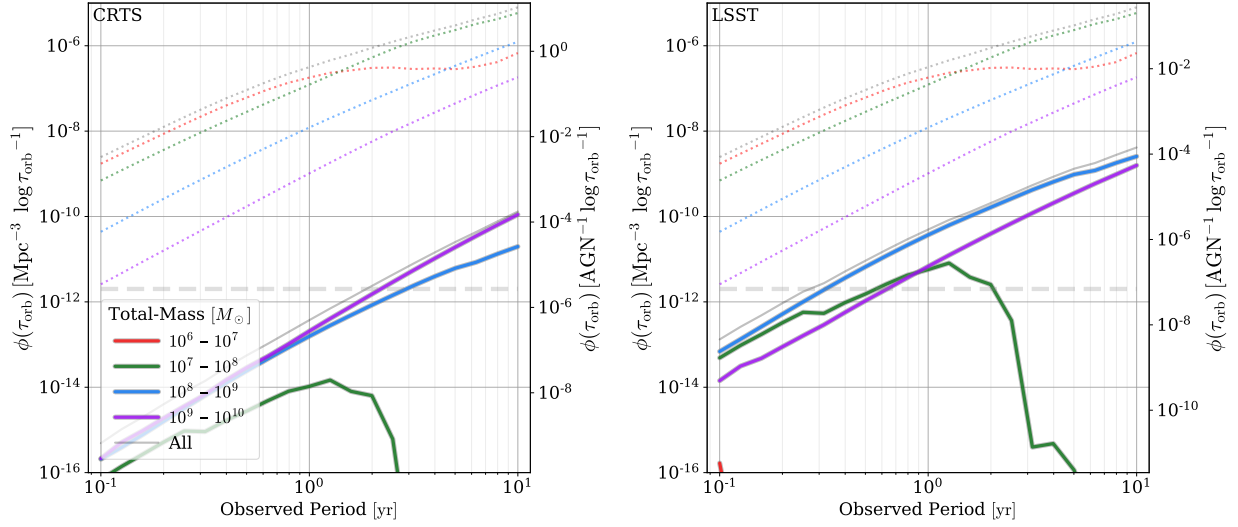
**Figure 4.5: Doppler-induced variability amplitudes** as a function of binary mass-ratio and total-mass. Blue colors denote regions where the secondary’s variations are dominant while red denotes that of the primary. An orbital period of  $\tau_{\text{orb}} = 1$  yr is assumed, and a system accretion-rate  $f_{\text{Edd,sys}} = 1$ . The accretion rate onto the secondary is shown in the green color-bar to the right. The hashed regions correspond to luminosities below the detection threshold for LSST (circles) and CRTS (dots) at redshift  $z = 1$ .

The secondary’s variation tends to dominate for systems with  $q \gtrsim 10^{-2}$ . When the primary dominates, the overall amplitude of variation is undetectably small ( $\lesssim 10^{-4}$ ). Systems with the largest variational amplitudes are concentrated around the largest masses, and mass-ratios near  $q \approx 0.1$ . While the region of parameter space with variations above  $\delta F_{\nu}^d \gtrsim 0.05$  is very small for  $f_{\text{Edd,sys}} = 1$ , i.e. only systems with  $M \gtrsim 2 \times 10^9 M_{\odot}$ , it increases for lower Eddington ratios. At  $f_{\text{Edd,sys}} = 10^{-1}$ , systems with  $M \gtrsim 2 \times 10^7 M_{\odot}$  can produce the same variability amplitudes.

The total luminosity of the system is tied closely to its total-mass. Near  $q \approx 0.1$ , the accretion-rate onto the secondary is significantly enhanced which decreased the luminosity of the primary—and thus the system overall. The hatched regions of Fig. 4.5 show systems with v-band spectral-fluxes below the LSST (circles) and CRTS (dots) thresholds for systems at a fixed luminosity distance of  $d_L(z = 1) \approx 6.5$  Gpc. At this distance, Eddington-limited systems can be seen down to  $M \approx 2 \times 10^8 M_{\odot}$  and  $M \approx 3 \times 10^6 M_{\odot}$  for CRTS and LSST like flux limits respectively.

#### 4.3.1.1 EVENT RATES

The population of observable, doppler-variable MBHB is shown in Fig. 4.6. The number-density of sources per decade of period is plotted for different total-mass bins. The dotted lines show the total number-densities of all binaries, while the solid lines show systems observable by CRTS-like (left-panel) and LSST-like (right-panel) surveys. Results are also shown in terms of occurrence rate relative to the



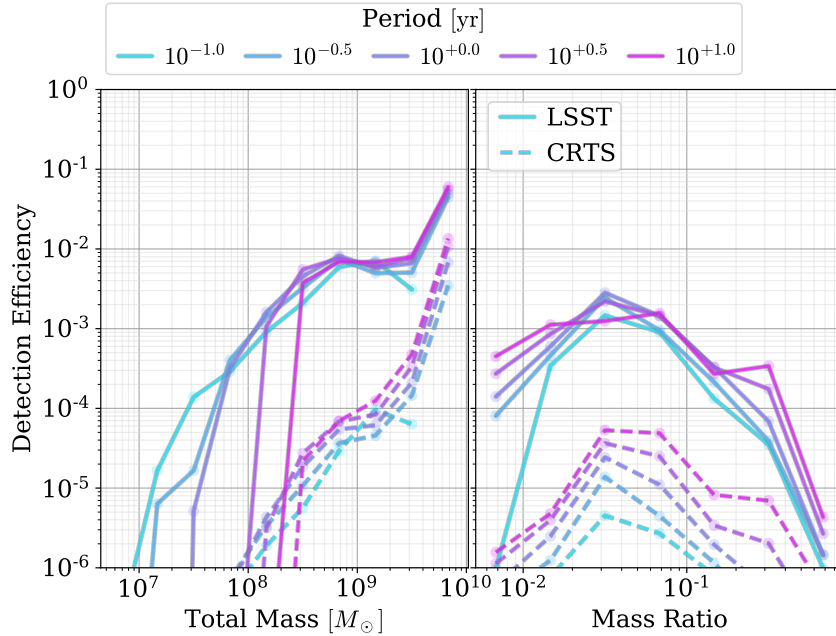
**Figure 4.6: Observability of Doppler-Variable MBH Binaries** versus period for CRTS (left) and LSST (right). The density of all binaries are shown with dotted lines, while observable systems are shown with solid lines. The densities are given per decade of binary period, i.e., the total number of binaries at any period is found by integrating these curves over  $d \log(\tau_{\text{orb}})$ . The left y-axes show systems in terms of volume densities, while the right y-axes give their number relative to that of all observable AGN for the same survey. The dashed, horizontal line indicated the number-density corresponding to a single, expected source to be detectable (i.e. the inverse of  $\approx 4 \times 10^5$  AGN expected to be observed by CRTS, and  $\approx 10^7$  by LSST).

number of observable AGN with the same instrument, which provides a more robust comparison to real observations.

The number-density of all binaries (dotted) is significantly higher for lower total-masses due to the strongly declining mass-function of MBH. In general, the density also declines sharply with decreasing orbital period, reflecting the GW-hardening timescale,  $\tau_{\text{gw}} \propto \tau_{\text{orb}}^{8/3}$ , by which nearly all binaries at these periods will be dominated\*. The systems which are observable as doppler-variables tend to be much more massive—for CRTS, predominantly above  $10^9 M_{\odot}$ . Observable systems tend to preserve a similar period dependence, with nearly two orders of magnitude fewer systems at  $\tau_{\text{var}} \approx 1$  yr than at  $\tau_{\text{var}} \approx 5$  yr. A future, more sensitive instrument like LSST can observe lower-mass systems: dominated by those between  $10^8 - 10^9 M_{\odot}$ . Even binaries between  $10^7 - 10^8 M_{\odot}$  are theoretically detectable at  $\tau_{\text{var}} \approx 1$  yr, but their occurrence rates may be too low to be seen. For comparison, the inverse of the total number of AGN expected to be detected by CRTS ( $\approx 4 \times 10^5$ ) and LSST ( $\approx 10^7$ ) is shown by the dashed, horizontal grey lines which give a sense of the minimum plausibly-detectable occurrence rates.

At periods between 1 – 5 yr, our models predict 9 doppler-variable MBHB to be observable by a CRTS-like survey, corresponding to  $2 \times 10^{-5} \text{ AGN}^{-1}$  ( $2 \times 10^{-11} \text{ arcsec}^{-2}$  out to redshift  $z_{\text{max}} = 2$ ). With LSST, this increases substantially to about 500 binaries, or  $3 \times 10^{-5} \text{ AGN}^{-1}$  ( $9 \times 10^{-10} \text{ arcsec}^{-2}$ ). While this detection rate is roughly an order of magnitude lower than the candidates from CRTS and PTF, it is encouraging that numerous real binaries are could be present in the data.

\*The plot of  $\phi$  for all MBHB in the  $10^6 - 10^7 M_{\odot}$  bin flattens out significantly at  $\tau_{\text{orb}} \gtrsim 1$  yr because environmental interactions, such as stellar-scattering, are still effective in this regime. Those interactions increase the hardening rate and thus decrease the number of binaries at those periods.



**Figure 4.7: Detection Efficiency of Doppler-Variable MBH Binaries** versus total mass (left) and mass ratio (right) for both LSST parameters (solid) and CRTS (dashed). Doppler-variable systems are only detectable when they have sufficiently large orbital velocities, leading to a strong drop-off in detection efficiency for lower-mass systems, especially at longer orbital periods. Lower mass-ratio systems tend to have larger velocities, but very low-mass secondaries become difficult to observe. This leads to a sweet-spot in mass-ratio for doppler-observability between  $q \approx 0.03 - 0.07$ .

#### 4.3.1.2 OBSERVATION EFFICIENCY

Figure 4.7 shows the efficiency of detection, i.e. the fraction of all systems which are observable, versus total mass (left) and mass ratio (right). Doppler-variable systems are only detectable when they have sufficiently large orbital velocities which leads to a strong drop-off in detection efficiency for lower-mass systems, especially at longer orbital periods. The highest efficiencies are achieved for the most massive systems, reaching  $\approx 10^{-2}$  for CRTS and almost  $10^{-1}$  for LSST. The total number of binaries is dominated by low-mass systems which are generally undetectable. Thus, at a given mass-ratio and marginalized over total masses, the efficiencies are extremely low:  $< 10^{-4}$  for CRTS and  $\lesssim 10^{-3}$  for LSST. There is a fairly well defined, efficiency sweet-spot for systems with mass-ratio between  $q \approx 0.03 - 0.07$ . This is determined by the interplay of lower-mass secondaries having larger velocities, but eventually becoming too faint to observe effectively.

The strong mass-dependency in detection efficiency strongly affects what binaries will be observed as doppler-variables in a given survey. This has important consequences for the implied GWB from a given set of observed systems. Consider a hypothetical set of doppler-variable binaries observed in an optical survey. The total-mass dependency implies that nearly all observed systems have  $M \gtrsim 3 \times 10^9 M_\odot$ , and less than  $\sim 1\%$  of all such binaries are observed. Furthermore, the mass-ratio dependence suggests that all of these systems have  $q \lesssim 0.1$ . Binaries with  $q \sim 1$  are roughly three times more common for periods of  $\sim 3$  yr (see Fig. 4.11), and each produces GW strains roughly ten times larger\*. Finally, recalling that strain amplitudes add in quadrature, this suggests that the GWB has a strain  $\approx 50$  times larger than the

\*Strain is related to chirp-mass as,  $h \propto \mathcal{M}^{5/3}$ , and the chirp-mass is  $\mathcal{M} \equiv Mq^{3/5}/(1+q)^{6/5}$ .

direct sum of the GW from the observed population itself.

### 4.3.2 HYDRODYNAMICALLY VARIABLE BINARIES

#### 4.3.2.1 EVENT RATES

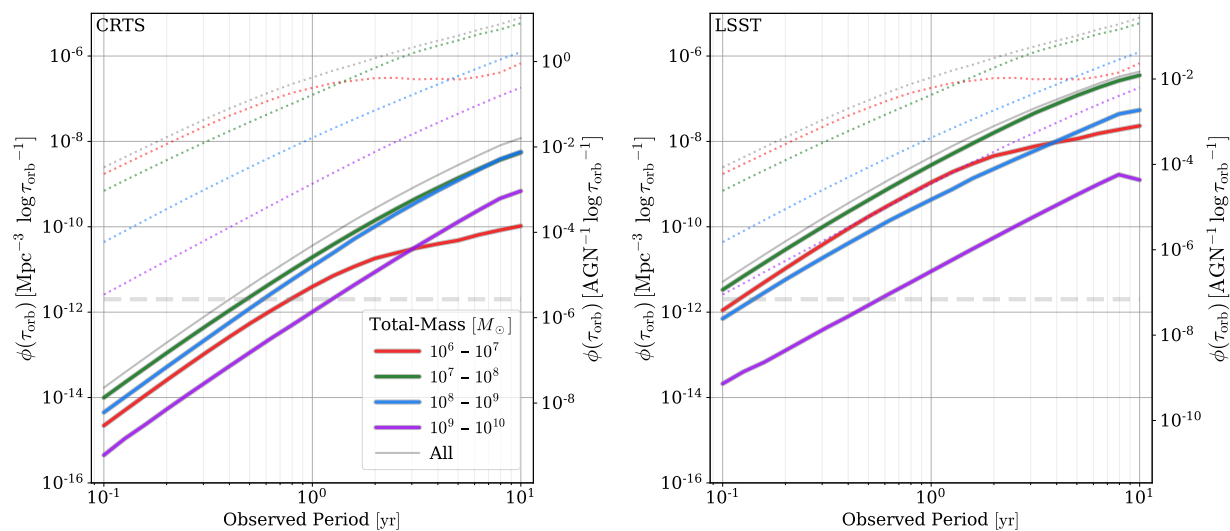
The population of observable, hydrodynamically variable MBHB is shown in Fig. 4.8. Overall, observabilities for hydrodynamic variables are typically two orders of magnitude higher than for doppler variables. Between 1 – 5 yr variability periods, for CRTS-like parameters, our model predicts roughly  $10^3$  systems to be observable, corresponding to  $3 \times 10^{-3} \text{ AGN}^{-1}$  ( $2 \times 10^{-9} \text{ arcsec}^{-2}$ ). This is notably higher than the occurrence rate of candidates identified in both CRTS and PTF, at rates of  $5 \times 10^{-4} \text{ AGN}^{-1}$  and  $10^{-3} \text{ AGN}^{-1}$  respectively. For an LSST-like survey, the predicted detection rate jumps to  $5 \times 10^4$  binaries, or  $4 \times 10^{-3} \text{ AGN}^{-1}$  ( $10^{-7} \text{ arcsec}^{-2}$ ).

The increased detection rates are driven primarily by the lack of explicit mass-dependence or period-dependence for hydrodynamic variability. The total-mass detection efficiency of doppler variables drops off rapidly for low mass systems, which dominate by number. Hydrodynamic variability only depends explicitly on mass-ratio, drastically increasing the available parameter space. This effect can be seen directly by comparing the CRTS rates between Fig. 4.6, and Fig. 4.8. In the latter, the detected systems between  $10^9$ – $10^{10} M_{\odot}$  increases by an order of magnitude, while that of  $10^8$ – $10^9 M_{\odot}$  systems increases by almost three orders. While doppler-variable  $10^7$ – $10^8 M_{\odot}$  systems are marginally observable near  $\tau_{\text{var}} \approx 1$  yr, the corresponding hydrodynamically variable systems are observable at rates comparable to  $10^8$ – $10^9 M_{\odot}$  at all periods.

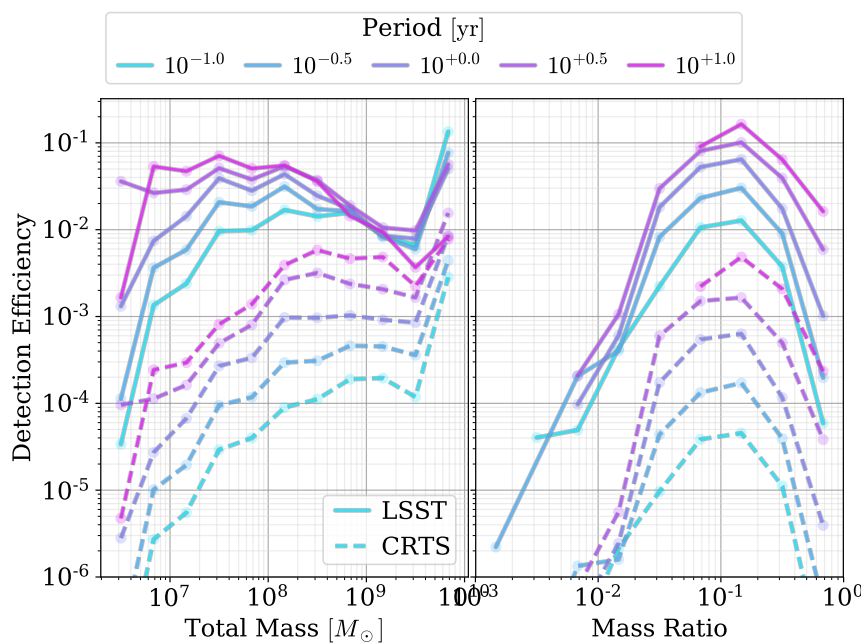
#### 4.3.2.2 OBSERVATION EFFICIENCY

Figure 4.9 shows the efficiency of detection versus total mass (left) and mass ratio (right) for hydrodynamically variable systems in different variability-period bins. Compared to doppler variability, the efficiency is relatively insensitive to total-mass, as mentioned above. With respect to both total-mass and mass-ratio, there is a strong trend towards lower detection efficiencies for shorter period systems. This seems to be driven primarily by disk truncation, which can significantly reduce the luminosity of each MBH’s disk. While the same effect is present for doppler variables, it is countered by the increasing orbital velocities at shorter periods.

The efficiency of hydrodynamic-variable detection with respect to mass-ratio is predominantly driven by two effects. First, for mass ratios  $q \ll 10^{-1}$ , the accretion rate to the secondary begins to drop off, leading to a strong decrease in detection efficiency. This effect is identically present for doppler-variables as well. Second, for mass ratios  $q > q_{\text{crit}} = 0.3$ , the variability timescale is shifted to  $\tau_{\text{var}} = 5 \tau_{\text{orb}}$ , which probes a much smaller population of systems at smaller orbital separations. This leads to a decrease in detections for high mass-ratio systems which mimics that seen in doppler-variables, although in that case due to lower secondary orbital-velocities.



**Figure 4.8: Observability of Hydrodynamic-Variable MBH Binaries** versus period for CRTS (left) and LSST (right). The total number of binaries are shown with dotted lines while the observable systems are shown with solid lines. Rates are given both in terms of number-density and relative to the number of observable AGN in the same survey.



**Figure 4.9: Detection Efficiency of Hydrodynamically-Variable MBH Binaries** versus total mass (left) and mass ratio (right) for both LSST parameters (solid) and CRTS (dashed).

## 4.4 DISCUSSION

In this paper, we make predictions for the electromagnetic detection of MBH binaries as variable AGN. We use a population of binaries drawn from cosmological hydrodynamic simulations of MBH and galaxies, evolved using semi-analytic, post-processing models of the detailed merger physics. Using synthetic spectra of AGN emission, along with models of both doppler and hydrodynamic variability, we have calculated detection rates for surveys with flux and variability sensitivities comparable to CRTS and LSST. Here we present the results and implications of our study, after first discussing their limitations.

### 4.4.1 CAVEATS

Numerous limitations exist in our current methods, both in terms of our binary populations and our models of variability. In the former class, while the masses of MBH in Illustris nicely reproduce the observed, redshift zero BH–galaxy scaling relations (Sijacki et al. 2015), there is still significant uncertainty in the full distribution of MBH masses (e.g. McConnell & Ma 2013). The MBH accretion rates have been calibrated to produce accurate masses and reproduce observational, bolometric luminosity functions (Sijacki et al. 2015). Using spectral models, and a simple model of obscuration, applied to the entire population of Illustris MBH, we predict a totally number of observable AGN (see Table 4.1) which differs by only 20% from CRTS observations (although we have not taken into account spectroscopic completeness). Still, the Illustris models were not designed with detailed spectral modeling in mind, and the 20% error surely underestimates systematic errors as seen in the discrepancies between our b-band luminosity-functions and observations (Fig. 4.4), for example.

When it comes to systems of high total-mass, our populations suffer from small-number statistics. As a method of reducing these effects, we have treated our binary population as averaged over redshift. In reality, the population of binaries at  $\sim$ yr periods are likely clustered near  $z \approx 1.0$  (see Fig. 4.10), which is not accounted for in our analysis, and similar to the known redshift distribution of AGN.

Instead of using bolometric luminosities and corrections or characteristic spectral indices, we have constructed full spectral models for each of the MBH in our binary populations. Still, these spectra are highly simplified in the complex and actively developing field of AGN emission. Perhaps the most important deficiency of our spectral models are the lack of any lines, extinction, or contributions from outside of the disk. We also do not consider any intrinsic AGN variability. Full spectroscopic observations of AGN, including variations between observing epoch, should be incorporated into calculations to carefully consider the effectiveness with which photometric variability can be accurately classified.

We have also relied very strongly on the results of Farris et al. (2014), which use 2D, isolated, purely-hydrodynamic simulations. Other groups have supported the Farris et al. (2014) results whose conclusions seem robust for their simulated conditions. Accounting for thick-disk accretion (and mass flow out of the disk plane, possibly enhanced by magnetic fields and radiation), and turbulent flows with varying inflow rates on large scales, are likely to change some of these results (at least quantitatively, if not qualitatively). In §4.4.2 we explore variations to many of our model parameters and assumptions, but our results seem quite robust.

Overall, we expect most of the simplifications and uncertainties in our models to tend towards *fewer* systems being observable as variables. Firstly, because biases in the  $M$ – $\sigma$  relation likely indicate lower overall MBH masses and luminosities. Secondly, thick-disk and turbulent accretion flows are more likely to smooth-out the periodic variations in emission rather than enhance them. Thirdly, AGN are known to exhibit strong intrinsic variability especially at low-frequencies which easily mimic periodicity. While



our model for variability sensitivity (Eq. 4.8) is based on observational studies of signal identification, it is extremely simplistic accounting for a very difficult task as shown in the careful analysis of [Graham et al. \(2015a\)](#), [Charisi et al. \(2016\)](#), and [Liu et al. \(2016\)](#). Finally, we believe it is likely that the accretion rates onto the binary MBH from Illustris are likely biased high. Taking these factors into account, we consider our predictions to be most robust as upper-bounds to expected detection rates.

#### 4.4.2 CONCLUSIONS

A summary of expected detection rates for variability periods between 1 – 5 yr are presented in Table 4.1. The sensitivities of current instruments like CRTS are likely to observe MBH binaries at roughly  $2 \times 10^{-5} \text{ AGN}^{-1}$  and  $3 \times 10^{-3} \text{ AGN}^{-1}$  for doppler and hydrodynamic variability, respectively, corresponding to roughly 9 and 1,000 sources in an all-sky survey. These expectations are an order of magnitude below and above the number of candidates put forward in CRTS by [Graham et al. 2015a](#) (which are consistent with the rate of candidates from [Charisi et al. 2016](#), and the upper limits from [Liu et al. 2016](#)). While there are reasons to expect that the published candidates contain false positives (e.g. [Sesana et al. 2017b](#)), our predictions indicate that there should exist a large number of true MBH binaries within the population.

Doppler variable sources are difficult to observe, primarily because near-relativistic orbital velocities are required to produce detectable oscillation amplitudes. Systems with high enough mass, and small enough orbital separations, tend to merge very quickly. Out of the published data sets, the large CRTS sample is most likely to contain doppler-variable binaries—up to  $\sim 10\%$  of the 111 candidate systems. Doppler variables should exhibit a characteristic spectral dependence (see Eq. 4.11 and [D’Orazio, Haiman & Schiminovich 2015](#)) which could be identified with continued spectroscopic, and perhaps even multi-band photometric, monitoring. Simply extending the observational baselines, while maintaining broadband temporal coverage, will of course also provide a powerful determinant in distinguishing red-noise contaminated systems.

Hydrodynamic variability is much less sensitive to binary mass and period, and thus probes a much broader parameter space replete with systems. Our models predict almost an order of magnitude more hydrodynamically variable systems to be observable than the number of published candidates. While a bias towards high accretion rates may cause a portion of this discrepancy, we believe it is likely that the induced, periodic hydrodynamic-variability is often drowned out by other aperiodic variations in the AGN luminosity. For example, disk turbulence and time-varying feeding rates of gas will not only introduce their own luminosity fluctuations, but also decrease the coherent, periodic variations seen in smooth, simulated accretion flows. Some simulations have also seen accretion alternate from primarily feeding one MBH to then predominantly feeding the other, even in otherwise smooth disks. This introduces a significant complication in separating AGN with significant red-noise from those which are binaries, but exhibit excursions from periodicity to due other disk phenomena.

Our models suggest that doppler-variable systems observable with current instruments should be entirely composed of binaries with total masses  $M_{\text{tot}} > 10^8 M_{\odot}$ , and dominated by systems with  $M_{\text{tot}} > 10^9 M_{\odot}$ . Even with a future instrument like LSST, few if any systems with  $M_{\text{tot}} < 10^8 M_{\odot}$  are likely to be observed. The situation is very different for hydrodynamic variables which, again, have no strong mass dependency on their variability amplitudes. Our models suggest that  $10^7 - 10^8 M_{\odot}$  systems and  $10^8 - 10^9 M_{\odot}$  systems should be observed with similar rates, while systems at  $M_{\text{tot}} > 10^9 M_{\odot}$  should be  $\approx 10$  times less common. Because of their relative rarity, doppler-variable systems are unlikely to be observed below  $\tau_{\text{var}} \approx 2 \text{ yr}$ , while hydrodynamically-variable systems could be seen at periods as short as

|                      | Observable, Variable Binaries |                    |                    |                    | AGN                |                    |
|----------------------|-------------------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
|                      | Doppler                       |                    | Hydrodynamic       |                    | CRTS               | LSST               |
|                      | CRTS                          | LSST               | CRTS               | LSST               |                    |                    |
| arcsec <sup>-2</sup> | $2 \times 10^{-11}$           | $1 \times 10^{-9}$ | $2 \times 10^{-9}$ | $1 \times 10^{-7}$ | $9 \times 10^{-7}$ | $3 \times 10^{-5}$ |
| number               | $9 \times 10^0$               | $5 \times 10^2$    | $1 \times 10^3$    | $5 \times 10^4$    | $4 \times 10^5$    | $1 \times 10^7$    |
| AGN <sup>-1</sup>    | $2 \times 10^{-5}$            | $3 \times 10^{-5}$ | $3 \times 10^{-3}$ | $4 \times 10^{-3}$ | -                  | -                  |

**Table 4.1: Expected Observability of MBH Binary, Periodically Variable-AGN** with periods between 1 and 5 years. Detection rates for surveys with parameters based on CRTS and LSST are shown, for doppler variability in the first two columns and hydrodynamic variability in the next two. The last two columns give the expected detection rates for (single) AGN. The first row gives detections per square-arcsecond, while the second row gives the total number of expected detections assuming a 33,000 deg<sup>2</sup> survey. For comparison, CRTS actually observes  $3.3 \times 10^5$  AGN (Graham et al. 2015a). The last row presents rates of binaries relative to the predicted number of AGN, which is likely less subject to systematic uncertainties in our luminosity functions and spectral models.

$$\tau_{\text{var}} \approx 0.3 \text{ yr.}$$

Both doppler and hydrodynamic variability strongly favor systems with mass-ratios  $q \sim 0.1$ , in our fiducial model. In the case of doppler variability, this trend is due to the larger orbital velocities of secondary MBH which have lower mass-ratios. For mass ratios  $q \lesssim 10^{-2}$ , the secondaries become too faint, and their variability signatures are harder to identify. The bias towards  $q \sim 0.1$  is enhanced by the heightened accretion rates for secondary MBH near that mass-ratio, which we impose based on the results of Farris et al. (2014). The accretion-partition function has the same effect on hydrodynamic variables. At the same time, binaries with  $q \approx 1$  are observed less effectively because their variability periods are shifted to much lower frequencies. Observing variability signatures at year periods are thus probing near equal-mass systems at much smaller orbital separations, where binaries merge more quickly, and thus where fewer systems are observable.

Sesana et al. (2017b) show that the CRTS candidates are in tension with current PTA upper limits on the presence of a GW background (GWB). Taking into account the parametric dependencies discussed above may increase that tension. In particular, Sesana et al. (2017b) show that if the intrinsic distribution of binary mass-ratios is biased towards lower values of  $q$ , those binaries produce weaker GW, and thus decrease tensions between the candidate populations and the PTA limits. Our binary populations have relatively flat mass-ratio distributions (with slight bias towards high  $q$ ), and the selection effects suggested by both our doppler and hydrodynamic variability models imply that the binaries which are *observed* tend to have lower mass-ratios. This observational bias has the opposite effect of a low- $q$  population, implying that high- $q$  systems, which dominate the GWB, are absent from the observational candidates, and would push the expected GWB amplitude higher. None the less, the total number of candidates identified (or excluded) by CRTS, PTF, and Pan-STARRS are all lower than our predicted rates of hydrodynamic variables, making the comparisons with the GWB estimates from our population non-trivial.

In conclusion, many MBH binaries should be present as periodically variable AGN in current optical surveys. Our understanding of the accretion rates onto year-period binaries, and the amount of intrinsic variability in accretion flows on similar scales is the largest uncertainty in understanding and identifying these systems. The work and candidate populations by Graham et al. (2015a), Charisi et al. (2016)

and [Liu et al. \(2016\)](#) present an incredible boon to AGN and MBH binary astrophysics. These candidates likely contain examples of sub-parsec MBHB which hitherto have never been confirmed. Such systems contain key information on MBH growth, MBH binary evolution, and offer stringent constraints and insights into our predictions for low-frequency GW signals soon to be detectable by PTA.

## SUPPLEMENTAL MATERIAL

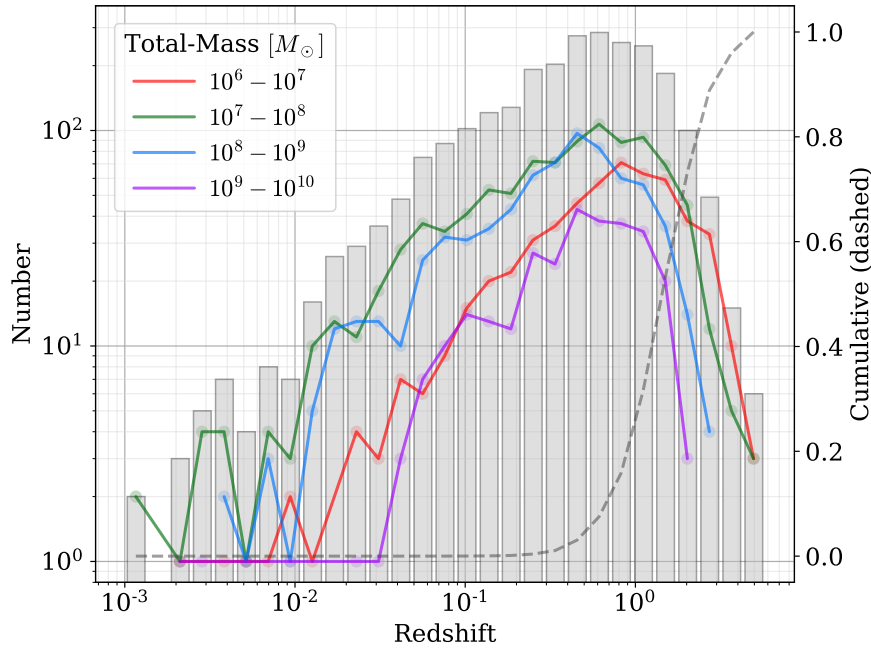
## ADDITIONAL FIGURES

Figure 4.10 shows the distribution of redshifts for Illustris binaries when they reach an orbital period of  $\tau_{\text{orb}} = 1$  yr. Out of  $\approx 30,000$  total binary mergers in Illustris, about 10,000 pass our selection cuts to assure both constituent host-galaxies are resolved (see, Kelley, Blecha & Hernquist 2017). Out of those, roughly 2,500 reach  $\tau_{\text{orb}} = 1$  yr before redshift zero, which are shown here. The grey histogram shows the distribution of all binaries while the colored lines show those of systems with the indicated total-masses and also with a mass ratio  $q \geq 10^{-2}$ . The mass-ratio cut only affects the systems with  $M_{\text{tot}} > 10^8 M_{\odot}$  (because the minimum MBH mass we consider is  $10^6 M_{\odot}$ ), and does not significantly effect their shape. While there is a noticeable tail of mergers down to redshifts  $z \lesssim 10^{-1}$ , the volume of the universe in this region is very small. To illustrate this, the distribution of binaries, weighted by comoving volume at each redshift bin, is shown with the dashed grey line which is entirely dominated by sources between  $z = 0.5 - 5.0$ .

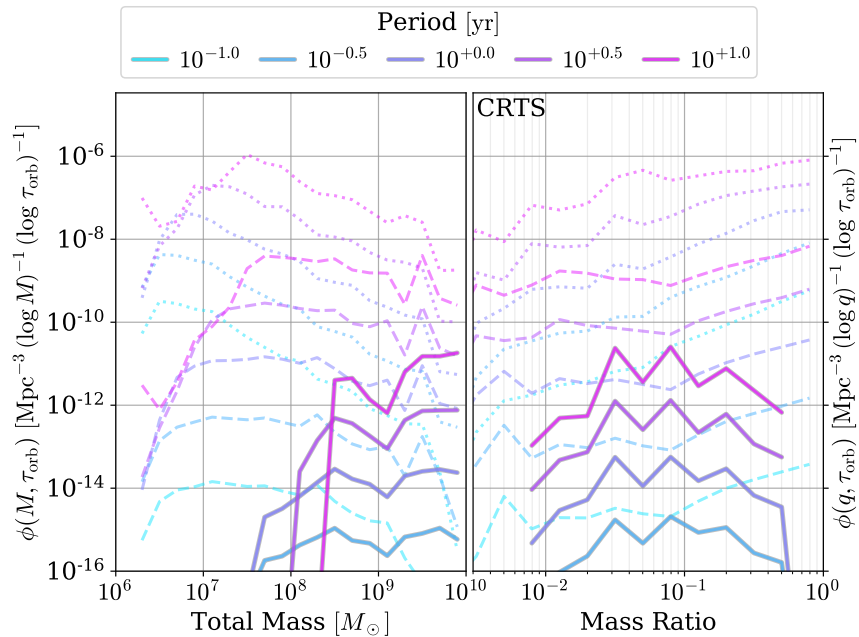
The parameters of all binaries (dashed) versus those which are observable as periodically-variable AGN (solid) are presented in the following figures. The left panel gives number-densities as a function of total mass, while the right gives them as a function of mass ratio. In each case a variety of orbital periods are shown (colors). The doppler-variable model is presented in Fig. 4.11 & Fig. 4.12 for CRTS & LSST respectively, while the hydrodynamic-variable model is presented in Fig. 4.13 & Fig. 4.14. Detection efficiencies—the ratios of observable- to all- systems—are presented in Figs. 4.7 & 4.9, doppler & hydrodynamic variability respectively. Knowing the total number of systems, however, is important for understanding selection effects, especially when inferring GWB amplitudes implied by an observed binary population (see §4.3.1.2 & §4.4).

For doppler variability (Figs. 4.11 & 4.12), there is a critical total-mass (left panels) at which the number of observed binaries drops rapidly for both CRTS and LSST, though for the latter the mass is much lower. This trend is a strong function of orbital period, as larger masses are required to obtain a given doppler-boost as periods increase. The observed number of sources drops significantly at high mass-ratios (right panels), while the actual number of systems with higher mass-ratios increases. This is due to two factors. First, the secondary—which tends to dominate the variability emission due to larger accretion rates—has smaller orbital velocities for higher mass-ratios. Second, the variability of the MBH is always opposite in phase, thus systems with more equal-masses tend to more effectively cancel out each-other’s luminosity variations.

Hydrodynamic variability (Figs. 4.13 & 4.14) is much less sensitive to the total mass of the binary (left panels), where the observed systems are fairly consistently proportional to all binaries. The mass-ratio dependence (right panels) however show a similar peak in sensitivity near  $q \approx 0.1$ . The decrease at higher mass-ratios is due primarily to our model assumption that the variational period is shifted to be a longer multiple of the orbital period for systems with high mass-ratios (see §4.2.4.1).



**Figure 4.10:** Redshift distribution of binaries reaching  $\tau_{\text{orb}} = 1$  yr, from the Illustris comoving volume of  $V_{\text{III}} = (108 \text{ Mpc})^3$ . The grey histogram shows the distribution of all systems, while the colored lines show those of binaries with total masses in the given range, and with mass ratios  $q \geq 10^{-2}$ . The grey dashed line shows the distribution of all binaries, but weighted by the comoving volume of the universe for the given redshift bin.



**Figure 4.11:** Doppler variable, CRTS-observable MBHB.

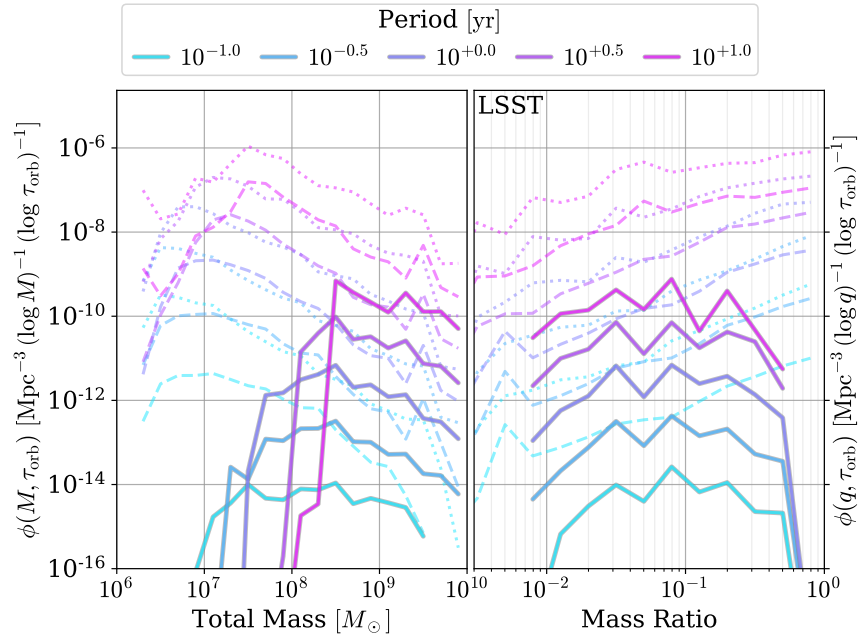


Figure 4.12: Doppler variable, LSST-observable MBHB.

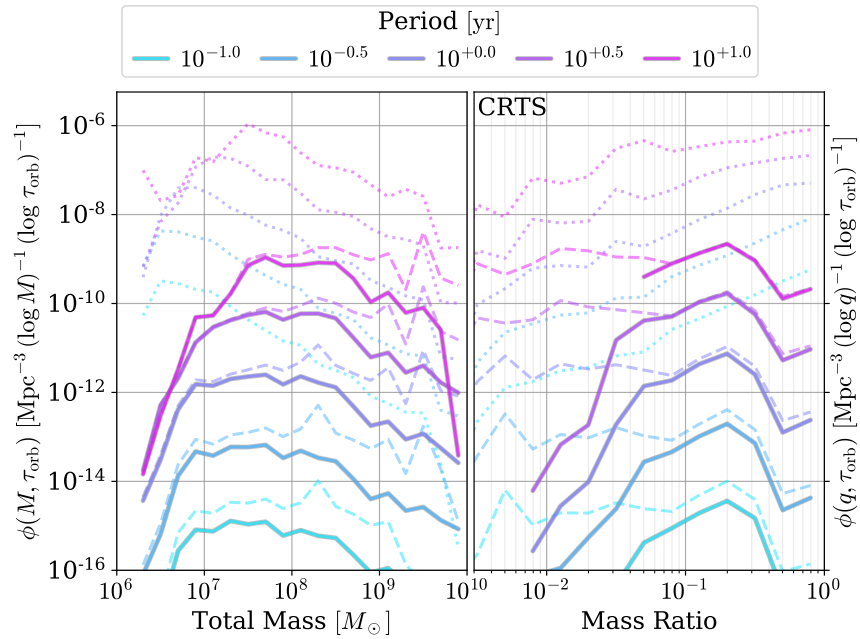


Figure 4.13: Hydrodynamically variable, CRTS-observable MBHB.

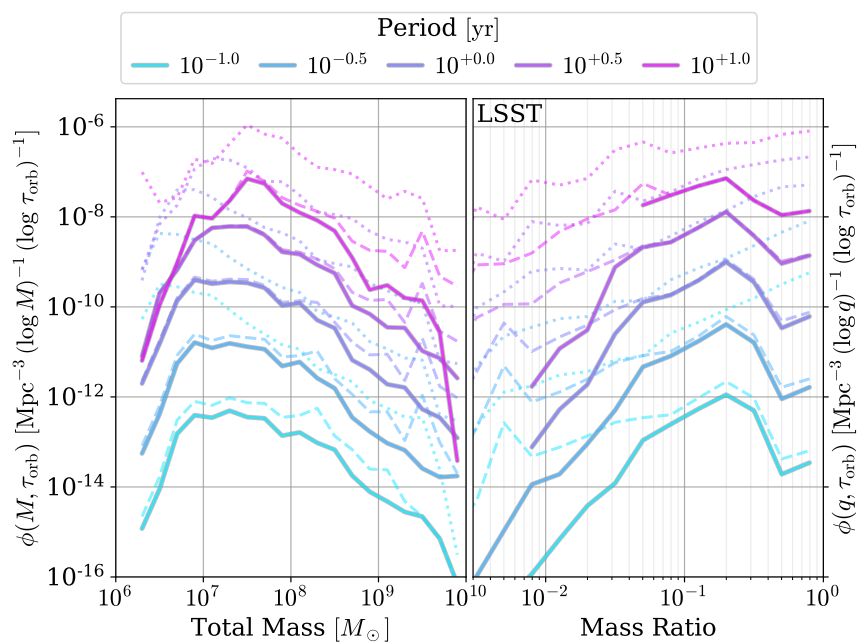


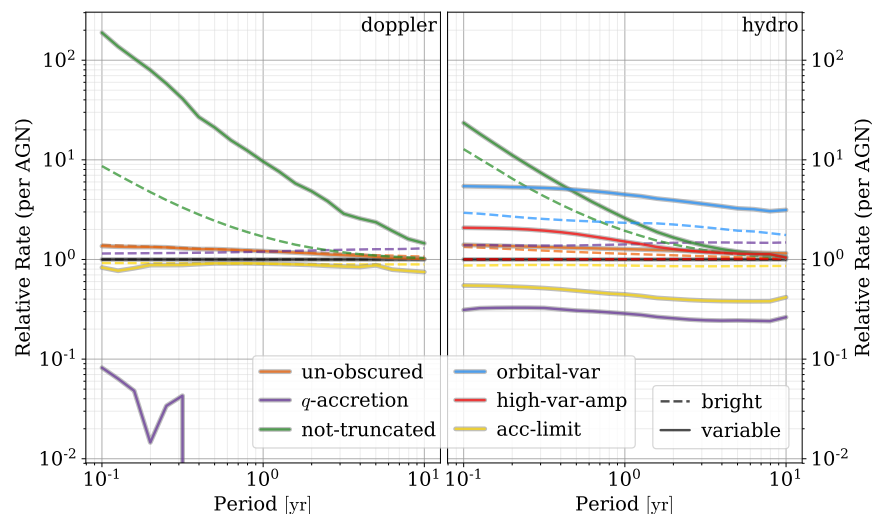
Figure 4.14: Hydrodynamically variable, LSST-observable MBHB.

## MODEL EFFECTS AND PARAMETERS

Our models of AGN variability include a variety of effects, and depend on many uncertain parameters. Figure 4.15 compares our fiducial detection rates with variations from different parametric changes to our models. Each line shows the ratio of detection rates for the varied model compared to that of the fiducial model, where detection rates are normalized to the expected number of observable AGN. The left panel shows doppler-variable binaries while the right shows hydrodynamically-variable ones, and in each case dashed lines show the ratio of systems which are bright enough to be observable (i.e. are above the CRTS flux limit), while solid lines show those with variability at a sufficient amplitude to also be detectable. The variations we consider are meant to illustrate both parametric uncertainties and the significance of particular physical effects. The alternative models shown are described below.

- Orange: neglecting AGN obscuration. Obscuration decreases the number of observably variable binaries by roughly a factor a four, but because the total number of observable AGN are decreased by nearly the same factor, the rate per AGN is nearly unaffected.
- Purple: scaling the accretion rate of each component MBH proportionally to its mass, i.e.  $\lambda \equiv \dot{M}_2/\dot{M}_1 = q$  (the fiducial model, based on Farris et al. 2014, is shown in Eq. 4.1 and Fig. 4.1). For both variability models, the secondary MBH is almost always the source of observable periodicity. Scaling the accretion rate of each component MBH to its mass produces significantly lower accretion rates to the secondary than using the Farris et al. 2014 model, thus significantly decreasing their observability as variable sources (solid lines). Because the total accretion rate to the system is unchanged, the number of sufficiently bright sources (dashed lines) are relatively unchanged, although slightly increased because more material is going to the brighter primary MBH. Doppler-variable sources are preferentially lower mass-ratio (those with larger secondary orbital velocities), and thus the observability of variable systems is almost entirely hindered, except for systems with extremely short orbital periods. For LSST, recall that our fiducial model predicts  $\sim 10^{-7}$  AGN $^{-1}$  binaries observable at these short periods (see Fig. 4.6), corresponding to order-unity doppler-variables detectable at  $\sim 0.3$  yr. If accretion is proportional to mass, the rate is decreased by another factor of  $\sim 20$ , making any detections very unlikely.
- Green: ignoring truncation of each MBH's accretion disk. Because each MBH is in a binary, its accretion disk can only extend to roughly their Hill spheres. The green curves show that this effect is very significant for both the overall brightness of sources and their detectable variability, especially at smaller binary separations (shorter periods). Using simple luminosity scaling arguments, which neglect disk truncation, will break down for periods below  $\sim 1$  yr.
- Blue: AGN variations at the orbital period. Some hydrodynamic simulations suggest that high mass-ratio systems will exhibit variability at periods roughly five times longer than the binary period, corresponding to the orbital time of an over-density of material at the outer edge of the gap in the disk. For these systems, observed periodicity at  $\sim 1$  yr is actually coming from sources with orbital periods of  $\sim 0.2$  yr, which are far less numerous. This hydrodynamic, frequency shift is not seen in all simulations. The model shown in blue assumes that the observed periodicity is always at the orbital period (i.e. unshifted), which significantly increases the number of observable systems by a factor of 3 – 5.
- Red: increased hydrodynamic variability-amplitude. In our fiducial model, we assume that the amplitude of accretion variability in the hydrodynamic model is  $\chi = 1.5$ . In the simulations of Farris et al. (2014), the authors find different variability amplitudes with different mass-ratios, gener-





**Figure 4.15:** Comparison of CRTS detection rates for parametric changes from our fiducial models. Each line shows the rate of observable-binaries per observable-AGN for the given model, divided by that of the fiducial model. The left panel shows doppler variables, while the right shows hydrodynamic ones. The following parametric variations are considered: (orange) ignoring obscuration of AGN emission; (purple) assuming that the accretion rate to each object is proportional to the mass ratio, i.e.  $\lambda = q$ ; (green) ignoring truncation of each MBH’s accretion disk; (blue) AGN variation occurs at the orbital period, regardless of mass ratio; (red) the hydrodynamic variability-amplitude ( $\chi$ ) is increased from 1.5 to 2.0; and (yellow) Eddington limited accretion for each MBH. Dashed lines show systems which are bright enough to be observable by CRTS, while the solid lines are systems which are both sufficiently bright and produce an observable amount of variability.

ally varying from  $\sim 1.5 - \sim 3.0$ . The model shown in red increases the variability amplitude to  $\chi = 2.0$ , which has little effect for long-period systems but up to a factor of two increase at short periods. Because long-period systems dominate, the overall change to detection rates is on the order of  $\sim 10\%$ . This is because the luminosity scales roughly with accretion rate, and even a 50% change in luminosity is easily observable in most systems. The effect increases at low frequencies because disk-truncation leads to more and more competition from the varying amplitude of the secondary MBH with the constant emission of the primary.

- Yellow: Eddington limit the accretion rate in each disk. Our fiducial model limits the combined accretion rate onto both MBH to the Eddington value. Because the accretion rate to the secondary, for moderate to large mass-ratios, is larger than the primary it can exceed the Eddington limit individually. The model shown with the yellow line assumes that both the combined and the individual accretion rates are both Eddington limited. Conceptually, this corresponds to effective regulation of the gas inflow even at small scales, or alternatively to the assumption that the radiative efficiency does not increase for Eddington fractions above unity\*. This model yields a small decrease to the number of doppler-variable sources,  $\approx 10\%$ ; and a more significant decrease to hydrodynamically-variable ones,  $\approx 60\%$ . The higher sensitivity to the accretion limit for hydrodynamically-variable systems is likely due to their sources tending to be at slightly higher mass-ratios ( $q \approx 0.1$ ) which is closer to the peak of the accretion partition function.

\*Some simulations have shown that the radiative efficiency increases only logarithmically for super-Eddington accretion rates in the ‘slim’-disk regime (e.g. Madau, Haardt & Dotti 2014).

Overall, uncertainty in how the accreting material is partitioned between MBH in the binary leads to the largest variations in our predicted populations. While our fiducial model of accretion which is enhanced for the secondary MBH produced many detectable doppler and hydrodynamically variable binaries by current surveys like CRTS. If the accretion is instead directly proportional to mass, then doppler variables are unlikely to ever be detected, even by LSST, and the number of hydrodynamic variables may be up to five times fewer. The relationship between orbital timescale and variational timescale also produces a significant effect, with our fiducial model predicting more conservative detection rates. We have also explored numerous other parametric variation to our models which were not included in Fig. 4.15, which all produce more minor effects. This is partially because we are normalizing our detection rates by the number of expected observable AGN. For example, doubling the sensitive flux of our modeled CRTS survey leads to a significant increase in the total number of predicted AGN and variables which are observable, but the variable rate is only changed by  $\sim 10\%$ .

# 5

## CONCLUSIONS

### 5.1 SUMMARY

In this thesis, we have made detailed predictions for the populations, lifetimes, and environments of MBH binaries using comprehensive models of the binary-merger process and state of the art, cosmological, hydrodynamic simulations. We have made calculated the prospective properties of both the GW background and individually-resolvable single-sources, along with expected times to detection using realistic models of pulsar timing arrays. In anticipation of the many multi-messenger opportunities between PTA and electromagnetic observations, we have explored the consistency between recently identified periodically-variable, MBHB candidates and the most stringent upper-limits on GW signals. We also generate predictions for the occurrence rates and parameter distributions of periodic-variable systems to better understand models of variability, the candidate populations, and to detection rates with future surveys.

In Ch. 1 (Kelley, Blecha & Hernquist 2017), we present our population of MBH binaries and their host galaxies derived from the Illustris cosmological hydrodynamic simulations. We detail our methods: using post-processing, semi-analytic models of the merger process to explore the parsec and sub-parsec scale dynamics which are unresolved in Illustris. The primary uncertainty in merger modeling is the effectiveness of refilling the stellar loss-cone, which largely determines what binary parameters can merge within a Hubble time. If refilling is inefficient, binaries with low total-masses ( $M \lesssim 10^8 M_\odot$ ) or with small mass-ratios ( $q \lesssim 10^{-2}$ ) are generally unable to coalesce before redshift zero. If loss-cone refilling is efficient, then only systems with both low total-masses *and* low mass-ratios stall. Typical redshifts of coalescence vary between  $z \sim 0.4 - 1.0$ , depending on hardening parameters. The number of MBH systems which are slow (or unable) to coalesce after the merger of their host galaxies suggest that three-body interactions from a subsequent galaxy merger could be important.

We use a semi-analytic model to calculate the GW Background amplitude from our binary populations. The predicted amplitudes are relatively insensitive to merger parameters, as the GWB is dominated by the most massive systems which, we find, tend to coalesce effectively. The maximum GWB amplitude from our population is  $A_{\text{yr}^{-1}} \approx 7 \times 10^{-16}$ , which is  $\approx 30\%$  below current PTA upper-limits, while more typical assumptions give amplitudes between  $A_{\text{yr}^{-1}} \approx 3-6 \times 10^{-16}$ . Stellar scattering, which increases the rate of binary inspiral at larger separations, decreases the amount of GW energy binaries can emit at lower frequencies. We find that this significantly flattens the GWB spectrum from the canonical  $-2/3$  strain spectral-index. For the most effective stellar scattering, this turnover can occur at frequencies as high as  $0.06 \text{ yr}^{-1} \approx 2 \text{ nHz}$ , which we do not expect to significantly hamper detection prospects with PTA.

Chapter 2 (Kelley et al. 2017) is dedicated to a careful calculation of the GWB, including numerous improvements to the methodology. In terms of dynamics, we present additions to our models to include eccentric binary evolution. In terms of the GWB calculation, we replace the semi-analytic models with Monte Carlo (MC) techniques—which better account for the finite population of binaries and for cosmic

variance. We show that eccentricity only significantly alters binary evolution when binaries are formed with highly elliptical orbits ( $e_0 \gtrsim 0.9$ ). Still, systems with initial eccentricities of  $e_0 = 0.5$  can enter the PTA band maintaining notably non-circular orbits, with  $e \approx 0.4$ . There are two primary effects of eccentricity on the GWB. First, non-zero eccentricity leads to more effective binary coalescence, which leads to increased attenuation of the emitted GW energy. Second, GW energy is shifted from lower to higher frequencies as individual binaries emit GW not only at twice the orbital period, but also at higher harmonics. For all but the highest initial eccentricities ( $e_0 \gtrsim 0.95$ ), non-circularity tends to slightly increase the GWB amplitude at PTA-sensitive frequencies. At high frequencies, eccentricity also remedies the steepening of the spectrum as the eccentric redistribution of GW energy leads to additional binaries contributing at higher observed frequencies. If the high-frequency spectrum can be measured in the future, it may be able to provide constraints on the binary-eccentricity distribution. At low frequencies, eccentricity tends to increase the strength of the GWB spectral turnover, but the location of the turnover still tends to be below the current PTA-band. In the future, observational constraints on the location of the turnover could provide insights into the typical properties of stellar environments around MBHB.

In this analysis, we also carefully examine plausible times-to-detection with parametrized models of each existing PTA. Based on our fiducial models, we find that all PTA should reach  $\sim 95\%$  detection probability between 2026 and 2034. If the International PTA is effective at combining data from the individual PTA and improving noise characterization, 95% detection probability could be reached as soon as 2024–2026. Times to detection vary by only a few years depending on hardening and eccentricity parameters which tends to be comparable to the variance between realizations.

In Ch. 3 (Kelley et al. 2018a), we turn our attention to signals from individually-resolvable, ‘deterministic’ sources which may be observable above the background. While previous studies have found that these single sources are likely to be much harder to detect than the GWB, we find instead that they should be detectable on comparable time-scales. At frequencies near  $0.1 \text{ yr}^{-1}$ , we find that there is roughly a 50% probability of a deterministic source producing timing residuals of  $\gtrsim 100 \text{ ns}$  (corresponding to characteristic strains  $h_c \approx 10^{-15}$ ). While individual binaries produce larger strains at higher frequencies, they also evolve much more quickly. This means that statistically, higher amplitude binaries are more common at lower frequencies. None the less, because the GWB decreases at higher frequencies, the occurrence rate of singles which are able to resound above the background increases at higher frequencies. Single-source properties are fairly insensitive to eccentricity, except for the largest values ( $e \gtrsim 0.95$ ).

We also present the first careful analysis of the effects of varying models for pulsar red-noise in estimated times to detection. Because red-noise dominates at low frequencies, as is expected from GW signals, it can be a significant contaminant in the data stream. Additionally, as the lowest accessible frequency bin depends on the total observing duration, red-noise can grow quickly with time. For these reasons, we find that detection prospects for the GWB are very sensitive to red-noise model, and unfortunately, the characterization of pulsar red-noise is remains highly uncertain. If all pulsars possess significant red-noise—which is, however, unlikely—it could prevent GWB detection altogether. More realistic assumptions for red-noise lead to delays in detection by a few years compared to white-noise only models which have typically been used in the literature. Unlike the GWB, single-source detection prospects are relatively insensitive to red-noise model as these deterministic sources are more common at higher frequencies.

Finally, in Ch. 4 (Kelley et al. 2018b), we predict the detection rates and parameter distributions of periodically-variable AGN. We combine our simulated MBHB populations with two leading models of producing flux variability: doppler-boosting, from the motion of systems with nearly relativistic orbital velocities; and hydrodynamic effects, which have been observed to produce periodically variable accretion rates in simulations of circumbinary disks. We find that, while doppler-variable systems should be present

in current survey data, they are likely too rare to explain the bulk of periodic-variability candidates. Observable doppler-variable systems are likely to have very large masses ( $\gtrsim 10^9 M_\odot$ ) with intermediate mass ratios ( $0.01 \lesssim q \lesssim 0.1$ ). The hydrodynamic-variability mechanism is much less sensitive to the total-mass of systems, which significantly increases the observable parameter space and thus their plausible detection rates. In fact, relative to the published variability candidates, our models for hydrodynamic variables over-predict the number of systems which should be observable. Likely the discrepancy is due to a combination of factors, including the accretion rates from Illustris being too high, and especially that realistic accretion flows onto each MBH will be more turbulent and intrinsically variable, both decreasing the sought after sinusoidal oscillations, and increasing the amount of intrinsic variability (i.e. noise) from which periodic signals must be extracted.

## 5.2 THE FUTURE

Many topics raised in this thesis are interesting for followup study, which we discuss here. Many projects can be approached with no, or only minor, additional modifications to the existing code and data sets. In particular: **(1)** calculating the electromagnetic observability of spatially offset (i.e. at  $\sim$ kpc separations; e.g. Comerford & Greene 2014) and kinematically offset (spectroscopically identifiable; e.g. Comerford et al. 2009a) binaries. In addition to the existing emission models we have constructed for accreting MBH *per se*, models for their local environments (particularly stellar cores) are also important for determining if the AGN will be identifiable. A related possibility is to consider the predicted observability of triple-AGN (e.g. Deane et al. 2014). Based on dual-AGN and galaxy-merger studies, we have a sense of the increased rate of fueling to AGN during galaxy interactions. Limits on the occurrence rates of triple-AGN systems could then be translated into constraints on the number of MBH stalled at kpc scales. **(2)** We have already explored GWB and GW single-sources, but we have yet to make predictions for GWB anisotropy (Mingarelli et al. 2013; Taylor & Gair 2013; Taylor et al. 2015), and for GW bursts-with-memory (Braginskii & Thorne 1987; Cordes & Jenet 2012; Arzoumanian et al. 2015), where, after MBHB coalescence, a permanent shift or ‘memory’ of the event remains and should be observable in pulsar timing. Our existing populations are sufficient to make predictions for both of these phenomena, simply additional analysis is required. **(3)** Calculating initial-eccentricity distributions of MBHB systems. The lack of accurate positional information from Illustris MBH (see §0.1.2) means that our initial eccentricity distributions had to be chosen arbitrarily. By tracking the nuclei of Illustris galaxies themselves, however, we can calculate initial orbital parameters in post-processing.

Some projects which would require the development of moderate, additional frameworks are also apparent. **(4)** Calculating the stellar tidal-disruption event (TDE) rate from MBHB, and the excavation of galactic cores. The TDE rate is believed to be enhanced by the presence of an MBH binary (e.g. Li et al. 2015, 2017), but this has never been calculated in a cosmological or MBHB population setting. Such a framework, would also lend itself towards examining whether the observed ‘cores’ in inner stellar-profiles can be explained by the ejection of stars during binary hardening, as has been proposed (Lauer et al. 2002, e.g.). **(5)** Incorporating three-body MBH interactions, and GW kicks into our merger simulations. Because the timescale for coalescence is often on the order of the Hubble time, subsequent galaxy mergers can bring in a third MBH. Recent studies have suggested that this could be an effective means of forcing some fraction of systems to coalesce regardless of environmental hardening mechanisms (Bonetti et al. 2017; Ryu et al. 2018). This is an important effect to consider in our existing calculations, as it could alter our GW predictions. Additionally, following coalescence, systems are expected to receive ‘kicks’ from the GW recoil (e.g. Blecha et al. 2016), which is also not included in our existing analyses.

There are also exciting avenues to pursue which would require significant developments, both in our existing methods, and requiring new tools and new simulations to be run. Specifically, developing high-resolution zoom-in simulations with accurate MBH dynamics and smaller seed-masses would provide a wealth of data for additional study. **(6)** Lower mass MBH would allow for predictions for LISA, which will be sensitive to binaries down to  $\sim 10^5 M_{\odot}$ , which are currently unresolved by Illustris. LISA will also be sensitive to much higher redshifts, necessitating MBH to be tracked at earlier stages in their evolution. **(7)** Resolving spatial scales can significantly improve our understanding of the MBHB merger dynamics. Using the Illustris data, we must extrapolate over many orders of magnitude to construct stellar and gas density profiles for the calculation of environmental interactions. Resolving these stars and gas down to  $\sim 1$  or even  $\sim 10$  pc would test the methods of these extrapolations, and increase the robustness of our calculations. Coupled with improvements to the MBH particle dynamics, the trajectories of MBH could also be followed much deeper into their evolution before semi-analytic models are required to take over. **(8)** Improved spatial resolution would also significantly improve our understanding of gas flows and AGN fueling during the merger process. This will be important for future infrared surveys like JWST and WFIRST which will better probe obscured galactic cores in the midst of merger. Higher resolution gas flows coupled to modern models for dust chemistry can also allow for drastically improved spectral models of AGN, both in general and for studies of dual and offset systems.

We have presented a variety of results making significant progress in our understanding of MBH binary populations and evolution. The implications for both GW and electromagnetic observations demonstrate an exciting future for multi-messenger observations of these systems in the next decade, and after. As is typically the case, our results have opened as many new questions. The projects outlined above represent numerous clear directions in which we can continue to progress the field. MBH binaries and their galactic environments offer a cornucopia of observations and physical phenomena to explore. I look forward to their continued study.



## APPENDIX

### 1.1 FREQUENTLY USED ABBREVIATIONS

|            |  |
|------------|--|
| ADAF       | Advection Dominated Accretion Flow                           |
| AGN        | Active Galactic Nuclei                                       |
| BH         | Black Hole   |
| CRTS       | Catalina Real-time Transient Survey                          |
| DF         | Dynamical Friction   |
| DP         | Detection Probability  |
| EM         | Electro-Magnetic   |
| EPTA       | European Pulsar Timing Array                                 |
| GW         | Gravitational Wave   |
| GWB        | Gravitational Wave Background                                |
| GWF        | Gravitational Wave Foreground                                |
| IPTA       | International Pulsar Timing Array                            |
| JWST       | James Webb Space Telescope                                   |
| LC         | (stellar) Loss-Cone  |
| LIGO       | Laser-Interferometer Gravitational-wave Observatory          |
| LISA       | Laser-Interferometer Space Antenna                           |
| LSST       | Large Synoptic Survey Telescope                              |
| MBH        | Massive Black-Hole   |
| MBHB       | Massive Black-Hole Binary                                    |
| NANOGrav   | North American Nanohertz Observatory for Gravitational Waves |
| Pan-STARRS | Panoramic Survey Telescope and Rapid Response System         |
| PPTA       | Parkes Pulsar Timing Array                                   |
| PTA        | Pulsar Timing Array  |
| PTF        | Palomar Transient Factory                                    |
| SMBH       | Super-Massive Black-Hole                                     |
| SNR        | Signal-to-Noise Ratio  |
| VD         | Viscous Drag   |
| WFIRST     | Wide-Field Infrared Survey Telescope                         |

## 1.2 FREQUENTLY USED SYMBOLS AND VALUES

| Symbol                        | Name                                     | Usage  |
|-------------------------------|--|--|
| $c$                           | Speed of Light                           | $3.0 \times 10^{10}$ cm/s                              |
| $G$                           | Newton's Gravitational Constant          | $6.7 \times 10^{-8}$ cm <sup>3</sup> g s <sup>-2</sup> |
| $M_{\odot}$                   | Solar Mass                               | $2.0 \times 10^{33}$ g                                 |
| $\varepsilon_{\text{rad}}$    | Radiative efficiency                     | Eq. 1  |
| $L_{\text{Edd}}$              | Eddington Luminosity                     | Eq. 2  |
| $\dot{M}_{\text{Edd}}$        | Eddington Accretion Rates                | Eq. 3  |
| $h_s$                         | GW Strain                                | Eq. 4  |
| $\mathcal{M}$                 | Chirp Mass                               | Eq. 5  |
| $A_{\text{yr}^{-1}}$          | GWB Amplitude at $f = 1 \text{ yr}^{-1}$ | Eq. 6  |
| $h_c$                         | Characteristic GW Strain                 | Eq. 9  |
| $\tau_{\text{dyn}}$           | Dynamical Times                          | Eq. 1.7  |
| $\mathcal{F}_{\text{refill}}$ | Loss-Cone Refilling Parameter            | Eq. 1.11   |
| $\mu, q$                      | Mass Ratio                               | $\mu \equiv M_2/M_1 \leq 1$                            |
| $f$                           | Observed frequency                       | $f = f_r (1 + z)$                                      |
| $f_r$                         | Rest-frame frequency                     | $f_r = f/(1 + z)$                                      |
| $R_s$                         | Schwarzschild Radius                     | $R_s \equiv 2GM/c^2$                                   |
| $d_L$                         | Luminosity Distance                      | $d_L = d_c (1 + z)$                                    |
| $d_c$                         | Comoving Distance                        | $d_c = d_L/(1 + z)$                                    |
| $f_{\text{Edd}}$              | Eddington Fraction                       | $f_{\text{Edd}} \equiv \dot{M}/\dot{M}_{\text{Edd}}$   |
| $T_{\text{obs}}$              | Total observing duration                 |  |
| $\Delta t_{\text{obs}}$       | Observing cadence                        |  |
| $\tau_{\text{var}}$           | Variability timescale                    |  |
| $\tau_{\text{orb}}$           | Orbital timescale                        |  |



## REFERENCES

- Amaro-Seoane, P.; Aoudia, S.; Babak, S.; Binétruy, P.; Berti, E.; Bohé, A.; Caprini, C.; Colpi, M.; Cornish, N.J.; Danzmann, K.; et al. (2012). Low-frequency gravitational-wave science with eLISA/NGO. *Classical and Quantum Gravity*, 29, 12, 124016. doi:[10.1088/0264-9381/29/12/124016](https://doi.org/10.1088/0264-9381/29/12/124016).
- Amaro-Seoane, P.; Gair, J.R.; Freitag, M.; Miller, M.C.; Mandel, I.; Cutler, C.J.; Babak, S. (2007). TOPICAL REVIEW: Intermediate and extreme mass-ratio inspirals—astrophysics, science applications and detection using LISA. *Classical and Quantum Gravity*, 24, R113–R169. doi:[10.1088/0264-9381/24/17/R01](https://doi.org/10.1088/0264-9381/24/17/R01).
- Amaro-Seoane, P.; Maureira-Fredes, C.; Dotti, M.; Colpi, M. (2016). Retrograde binaries of massive black holes in circumbinary accretion discs. *Astron. Astrophys.*, 591, A114. doi:[10.1051/0004-6361/201526172](https://doi.org/10.1051/0004-6361/201526172).
- Amaro-Seoane, P.; Sesana, A.; Hoffman, L.; Benacquista, M.; Eichhorn, C.; Makino, J.; Spurzem, R. (2010). Triplets of supermassive black holes: astrophysics, gravitational waves and detection. *Mon. Not. R. Astron. Soc.*, 402, 2308–2320. doi:[10.1111/j.1365-2966.2009.16104.x](https://doi.org/10.1111/j.1365-2966.2009.16104.x).
- Anholm, M.; Ballmer, S.; Creighton, J.D.E.; Price, L.R.; Siemens, X. (2009). Optimal strategies for gravitational wave stochastic background searches in pulsar timing data. *Phys. Rev. D*, 79, 8, 084030. doi:[10.1103/PhysRevD.79.084030](https://doi.org/10.1103/PhysRevD.79.084030).
- Antonini, F.; Merritt, D. (2012). Dynamical Friction around Supermassive Black Holes. *Astrophys. J.*, 745, 83. doi:[10.1088/0004-637X/745/1/83](https://doi.org/10.1088/0004-637X/745/1/83).
- Antonucci, R. (1993). Unified models for active galactic nuclei and quasars. *Annu. Rev. Astron. Astrophys.*, 31, 473–521. doi:[10.1146/annurev.aa.31.090193.002353](https://doi.org/10.1146/annurev.aa.31.090193.002353).
- Armano, M.; Audley, H.; Auger, G.; Baird, J.T.; Bassan, M.; Binetruy, P.; Born, M.; Bortoluzzi, D.; Brandt, N.; Caleno, M.; et al. (2016). Sub-Femto-g Free Fall for Space-Based Gravitational Wave Observatories: LISA Pathfinder Results. *Physical Review Letters*, 116, 23, 231101. doi:[10.1103/PhysRevLett.116.231101](https://doi.org/10.1103/PhysRevLett.116.231101).
- Armitage, P.J.; Natarajan, P. (2005). Eccentricity of Supermassive Black Hole Binaries Coalescing from Gas-rich Mergers. *Astrophys. J.*, 634, 921–927. doi:[10.1086/497108](https://doi.org/10.1086/497108).
- Arzoumanian, Z.; Brazier, A.; Burke-Spolaor, S.; Chamberlin, S.J.; Chatterjee, S.; Christy, B.; Cordes, J.M.; Cornish, N.J.; Crowter, K.; Demorest, P.B.; et al. (2016). The NANOGrav Nine-year Data Set: Limits on the Isotropic Stochastic Gravitational Wave Background. *Astrophys. J.*, 821, 13. doi:[10.3847/0004-637X/821/1/13](https://doi.org/10.3847/0004-637X/821/1/13).
- Arzoumanian, Z.; Brazier, A.; Burke-Spolaor, S.; Chamberlin, S.J.; Chatterjee, S.; Christy, B.; Cordes, J.M.; Demorest, P.B.; Deng, X.; et al. (2015). NANOGrav Constraints on Gravitational Wave Bursts with Memory. *Astrophys. J.*, 810, 150. doi:[10.1088/0004-637X/810/2/150](https://doi.org/10.1088/0004-637X/810/2/150).
- Arzoumanian, Z.; Brazier, A.; Burke-Spolaor, S.; Chamberlin, S.J.; Chatterjee, S.; Cordes, J.M.; Demorest, P.B.; Deng, X.; Dolch, T.; Ellis, J.A.; et al. (2014). Gravitational Waves from Individual Supermassive Black Hole Binaries in Circular Orbits: Limits from the North American Nanohertz Observatory for Gravitational Waves. *Astrophys. J.*, 794, 141. doi:[10.1088/0004-637X/794/2/141](https://doi.org/10.1088/0004-637X/794/2/141).
- Astropy Collaboration; Robitaille, T.P.; Tollerud, E.J.; Greenfield, P.; Droettboom, M.; Bray, E.; Aldcroft, T.; Davis, M.; Ginsburg, A.; Price-Whelan, A.M.; et al. (2013). Astropy: A community Python pack-

age for astronomy. *Astron. Astrophys.*, 558, A33. doi:[10.1051/0004-6361/201322068](https://doi.org/10.1051/0004-6361/201322068).

Babak, S.; Petiteau, A.; Sesana, A.; Brem, P.; Rosado, P.A.; Taylor, S.R.; Lassus, A.; Hessels, J.W.T.; Bassa, C.G.; Burgay, M.; et al. (2016). European Pulsar Timing Array limits on continuous gravitational waves from individual supermassive black hole binaries. *Mon. Not. R. Astron. Soc.*, 455, 1665–1679. doi:[10.1093/mnras/stv2092](https://doi.org/10.1093/mnras/stv2092).

Babak, S.; Sesana, A. (2012). Resolving multiple supermassive black hole binaries with pulsar timing arrays. *Phys. Rev. D*, 85, 4, 044034. doi:[10.1103/PhysRevD.85.044034](https://doi.org/10.1103/PhysRevD.85.044034).

Balbus, S.A.; Hawley, J.F. (1998). Instability, turbulence, and enhanced transport in accretion disks. *Reviews of Modern Physics*, 70, 1–53. doi:[10.1103/RevModPhys.70.1](https://doi.org/10.1103/RevModPhys.70.1).

Barnes, J.; Hut, P. (1986). A hierarchical  $O(N \log N)$  force-calculation algorithm. *Nature*, 324, 446–449. doi:[10.1038/324446a0](https://doi.org/10.1038/324446a0).

Barnes, J.E.; Hernquist, L. (1992). Dynamics of interacting galaxies. *Annu. Rev. Astron. Astrophys.*, 30, 705–742. doi:[10.1146/annurev.aa.30.090192.003421](https://doi.org/10.1146/annurev.aa.30.090192.003421).

Barrows, R.S.; Comerford, J.M.; Greene, J.E.; Pooley, D. (2016). Spatially Offset Active Galactic Nuclei. I. Selection and Spectroscopic Properties. *Astrophys. J.*, 829, 37. doi:[10.3847/0004-637X/829/1/37](https://doi.org/10.3847/0004-637X/829/1/37).

Barrows, R.S.; Comerford, J.M.; Greene, J.E.; Pooley, D. (2017). Spatially Offset Active Galactic Nuclei. II. Triggering in Galaxy Mergers. *Astrophys. J.*, 838, 129. doi:[10.3847/1538-4357/aa64d9](https://doi.org/10.3847/1538-4357/aa64d9).

Begelman, M.C.; Blandford, R.D.; Rees, M.J. (1980). Massive black hole binaries in active galactic nuclei. *Nature*, 287, 307–309. doi:[10.1038/287307a0](https://doi.org/10.1038/287307a0).

Begelman, M.C.; Volonteri, M.; Rees, M.J. (2006). Formation of supermassive black holes by direct collapse in pre-galactic haloes. *Mon. Not. R. Astron. Soc.*, 370, 289–298. doi:[10.1111/j.1365-2966.2006.10467.x](https://doi.org/10.1111/j.1365-2966.2006.10467.x).

Berczik, P.; Merritt, D.; Spurzem, R.; Bischof, H.P. (2006). Efficient Merger of Binary Supermassive Black Holes in Nonaxisymmetric Galaxies. *Astrophys. J. Lett.*, 642, L21–L24. doi:[10.1086/504426](https://doi.org/10.1086/504426).

Bertotti, B.; Carr, B.J.; Rees, M.J. (1983). Limits from the timing of pulsars on the cosmic gravitational wave background. *Mon. Not. R. Astron. Soc.*, 203, 945–954. doi:[10.1093/mnras/203.4.945](https://doi.org/10.1093/mnras/203.4.945).

Binney, J.; Tremaine, S. (1987). *Galactic dynamics*. Princeton University Press.

Blaes, O.; Lee, M.H.; Socrates, A. (2002). The Kozai Mechanism and the Evolution of Binary Supermassive Black Holes. *Astrophys. J.*, 578, 775–786. doi:[10.1086/342655](https://doi.org/10.1086/342655).

Blandford, R.; Romani, R.W.; Narayan, R. (1984). Arrival-time analysis for a millisecond pulsar. *Journal of Astrophysics and Astronomy*, 5, 369–388. doi:[10.1007/BF02714466](https://doi.org/10.1007/BF02714466).

Blecha, L.; Cox, T.J.; Loeb, A.; Hernquist, L. (2011). Recoiling black holes in merging galaxies: relationship to active galactic nucleus lifetimes, starbursts and the  $M_{BH}-\sigma_*$  relation. *Mon. Not. R. Astron. Soc.*, 412, 2154–2182. doi:[10.1111/j.1365-2966.2010.18042.x](https://doi.org/10.1111/j.1365-2966.2010.18042.x).

Blecha, L.; Sijacki, D.; Kelley, L.Z.; Torrey, P.; Vogelsberger, M.; Nelson, D.; Springel, V.; Snyder, G.;

- Hernquist, L. (2016). Recoiling black holes: prospects for detection and implications of spin alignment. *Mon. Not. R. Astron. Soc.*, 456, 961–989. doi:[10.1093/mnras/stv2646](https://doi.org/10.1093/mnras/stv2646).
- Blumenthal, G.R.; Faber, S.M.; Primack, J.R.; Rees, M.J. (1984). Formation of galaxies and large-scale structure with cold dark matter. *Nature*, 311, 517–525. doi:[10.1038/311517a0](https://doi.org/10.1038/311517a0).
- Bogdanović, T.; Reynolds, C.S.; Miller, M.C. (2007). Alignment of the Spins of Supermassive Black Holes Prior to Coalescence. *Astrophys. J. Lett.*, 661, L147–L150. doi:[10.1086/518769](https://doi.org/10.1086/518769).
- Bondi, H. (1952). On spherically symmetrical accretion. *Mon. Not. R. Astron. Soc.*, 112, 195. doi:[10.1093/mnras/112.2.195](https://doi.org/10.1093/mnras/112.2.195).
- Bonetti, M.; Haardt, F.; Sesana, A.; Barausse, E. (2016). Post-Newtonian evolution of massive black hole triplets in galactic nuclei - I. Numerical implementation and tests. *Mon. Not. R. Astron. Soc.*, 461, 4419–4434. doi:[10.1093/mnras/stw1590](https://doi.org/10.1093/mnras/stw1590).
- Bonetti, M.; Sesana, A.; Barausse, E.; Haardt, F. (2017). Post-Newtonian evolution of massive black hole triplets in galactic nuclei – III. A robust lower limit to the nHz stochastic background of gravitational waves. *ArXiv e-prints*.
- Boroson, T.A.; Lauer, T.R. (2009). A candidate sub-parsec supermassive binary black hole system. *Nature*, 458, 53–55. doi:[10.1038/nature07779](https://doi.org/10.1038/nature07779).
- Boyle, L.; Pen, U.L. (2012). Pulsar timing arrays as imaging gravitational wave telescopes: Angular resolution and source (de)confusion. *Phys. Rev. D*, 86, 12, 124028. doi:[10.1103/PhysRevD.86.124028](https://doi.org/10.1103/PhysRevD.86.124028).
- Braginskii, V.B.; Thorne, K.S. (1987). Gravitational-wave bursts with memory and experimental prospects. *Nature*, 327, 123–125. doi:[10.1038/327123a0](https://doi.org/10.1038/327123a0).
- Bromm, V.; Loeb, A. (2003). Formation of the First Supermassive Black Holes. *Astrophys. J.*, 596, 34–46. doi:[10.1086/377529](https://doi.org/10.1086/377529).
- Burke-Spolaor, S. (2011). A radio Census of binary supermassive black holes. *Mon. Not. R. Astron. Soc.*, 410, 2113–2122. doi:[10.1111/j.1365-2966.2010.17586.x](https://doi.org/10.1111/j.1365-2966.2010.17586.x).
- Burke-Spolaor, S. (2013). Multi-messenger approaches to binary supermassive black holes in the ‘continuous-wave’ regime. *Classical and Quantum Gravity*, 30, 22, 224013. doi:[10.1088/0264-9381/30/22/224013](https://doi.org/10.1088/0264-9381/30/22/224013).
- Caballero, R.N.; Lee, K.J.; Lentati, L.; Desvignes, G.; Champion, D.J.; Verbiest, J.P.W.; Janssen, G.H.; Stappers, B.W.; Kramer, M.; Lazarus, P.; et al. (2016). The noise properties of 42 millisecond pulsars from the European Pulsar Timing Array and their impact on gravitational-wave searches. *Mon. Not. R. Astron. Soc.*, 457, 4421–4440. doi:[10.1093/mnras/stw179](https://doi.org/10.1093/mnras/stw179).
- Cen, R. (1992). A hydrodynamic approach to cosmology - Methodology. *Astrophys. J., Suppl. Ser.*, 78, 341–364. doi:[10.1086/191630](https://doi.org/10.1086/191630).
- Chandrasekhar, S. (1942). *Principles of stellar dynamics*.
- Chandrasekhar, S. (1943). Dynamical Friction. I. General Considerations: the Coefficient of Dynamical Friction. *Astrophys. J.*, 97, 255. doi:[10.1086/144517](https://doi.org/10.1086/144517).

- Charisi, M.; Bartos, I.; Haiman, Z.; Price-Whelan, A.M.; Graham, M.J.; Bellm, E.C.; Laher, R.R.; Márka, S. (2016). A population of short-period variable quasars from PTF as supermassive black hole binary candidates. *Mon. Not. R. Astron. Soc.*, 463, 2145–2171. doi:[10.1093/mnras/stw1838](https://doi.org/10.1093/mnras/stw1838).
- Charisi, M.; Haiman, Z.; Schiminovich, D.; D’Orazio, D.J. (2018). Testing the relativistic Doppler boost hypothesis for supermassive black hole binary candidates. *Mon. Not. R. Astron. Soc.*. doi:[10.1093/mnras/sty516](https://doi.org/10.1093/mnras/sty516).
- Chen, S.; Middleton, H.; Sesana, A.; Del Pozzo, W.; Vecchio, A. (2017). Probing the assembly history and dynamical evolution of massive black hole binaries with pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 468, 404–417. doi:[10.1093/mnras/stx475](https://doi.org/10.1093/mnras/stx475).
- Chen, S.; Sesana, A.; Del Pozzo, W. (2016). Efficient computation of the gravitational wave spectrum emitted by eccentric massive black hole binaries in stellar environments. *ArXiv e-prints*.
- Chen, X.; Madau, P.; Sesana, A.; Liu, F.K. (2009). Enhanced Tidal Disruption Rates from Massive Black Hole Binaries. *Astrophys. J. Lett.*, 697, L149–L152. doi:[10.1088/0004-637X/697/2/L149](https://doi.org/10.1088/0004-637X/697/2/L149).
- Colella, P.; Woodward, P.R. (1984). The Piecewise Parabolic Method (PPM) for Gas-Dynamical Simulations. *Journal of Computational Physics*, 54, 174–201. doi:[10.1016/0021-9991\(84\)90143-8](https://doi.org/10.1016/0021-9991(84)90143-8).
- Colpi, M.; Mayer, L.; Governato, F. (1999). Dynamical Friction and the Evolution of Satellites in Virialized Halos: The Theory of Linear Response. *Astrophys. J.*, 525, 720–733. doi:[10.1086/307952](https://doi.org/10.1086/307952).
- Comerford, J.M.; Gerke, B.F.; Newman, J.A.; Davis, M.; Yan, R.; Cooper, M.C.; Faber, S.M.; Koo, D.C.; Coil, A.L.; Rosario, D.J.; et al. (2009a). Inspiralling Supermassive Black Holes: A New Signpost for Galaxy Mergers. *Astrophys. J.*, 698, 956–965. doi:[10.1088/0004-637X/698/1/956](https://doi.org/10.1088/0004-637X/698/1/956).
- Comerford, J.M.; Gerke, B.F.; Stern, D.; Cooper, M.C.; Weiner, B.J.; Newman, J.A.; Madsen, K.; Barrows, R.S. (2012). Kiloparsec-scale Spatial Offsets in Double-peaked Narrow-line Active Galactic Nuclei. I. Markers for Selection of Compelling Dual Active Galactic Nucleus Candidates. *Astrophys. J.*, 753, 42. doi:[10.1088/0004-637X/753/1/42](https://doi.org/10.1088/0004-637X/753/1/42).
- Comerford, J.M.; Greene, J.E. (2014). Offset Active Galactic Nuclei as Tracers of Galaxy Mergers and Supermassive Black Hole Growth. *Astrophys. J.*, 789, 112. doi:[10.1088/0004-637X/789/2/112](https://doi.org/10.1088/0004-637X/789/2/112).
- Comerford, J.M.; Griffith, R.L.; Gerke, B.F.; Cooper, M.C.; Newman, J.A.; Davis, M.; Stern, D. (2009b). 1.75 h<sup>-1</sup> kpc Separation Dual Active Galactic Nuclei at z = 0.36 in the Cosmos Field. *Astrophys. J. Lett.*, 702, L82–L86. doi:[10.1088/0004-637X/702/1/L82](https://doi.org/10.1088/0004-637X/702/1/L82).
- Comerford, J.M.; Pooley, D.; Barrows, R.S.; Greene, J.E.; Zakamska, N.L.; Madejski, G.M.; Cooper, M.C. (2015). Merger-driven Fueling of Active Galactic Nuclei: Six Dual and Offset AGNs Discovered with Chandra and Hubble Space Telescope Observations. *Astrophys. J.*, 806, 219. doi:[10.1088/0004-637X/806/2/219](https://doi.org/10.1088/0004-637X/806/2/219).
- Corbin, V.; Cornish, N.J. (2010). Pulsar Timing Array Observations of Massive Black Hole Binaries. *ArXiv e-prints*.
- Cordes, J.M.; Jenet, F.A. (2012). Detecting Gravitational Wave Memory with Pulsar Timing. *Astrophys. J.*, 752, 54. doi:[10.1088/0004-637X/752/1/54](https://doi.org/10.1088/0004-637X/752/1/54).

- Cornish, N.J.; Sampson, L. (2016). Towards robust gravitational wave detection with pulsar timing arrays. *Phys. Rev. D*, 93, 10, 104047. doi:[10.1103/PhysRevD.93.104047](https://doi.org/10.1103/PhysRevD.93.104047).
- Courant, R.; Friedrichs, K.O. (1948). *Supersonic flow and shock waves*.
- Cuadra, J.; Armitage, P.J.; Alexander, R.D.; Begelman, M.C. (2009a). Massive black hole binary mergers within subparsec scale gas discs. *Mon. Not. R. Astron. Soc.*, 393, 1423–1432. doi:[10.1111/j.1365-2966.2008.14147.x](https://doi.org/10.1111/j.1365-2966.2008.14147.x).
- Cuadra, J.; Armitage, P.J.; Alexander, R.D.; Begelman, M.C. (2009b). Massive black hole binary mergers within subparsec scale gas discs. *Mon. Not. R. Astron. Soc.*, 393, 1423–1432. doi:[10.1111/j.1365-2966.2008.14147.x](https://doi.org/10.1111/j.1365-2966.2008.14147.x).
- Davis, M.; Efstathiou, G.; Frenk, C.S.; White, S.D.M. (1985). The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.*, 292, 371–394. doi:[10.1086/163168](https://doi.org/10.1086/163168).
- Davis, S.W.; Laor, A. (2011). The Radiative Efficiency of Accretion Flows in Individual Active Galactic Nuclei. *Astrophys. J.*, 728, 98. doi:[10.1088/0004-637X/728/2/98](https://doi.org/10.1088/0004-637X/728/2/98).
- De Paolis, F.; Ingrosso, G.; Nucita, A.A. (2002). Astrophysical implications of binary black holes in BL Lacertae objects. *Astron. Astrophys.*, 388, 470–476. doi:[10.1051/0004-6361:20020519](https://doi.org/10.1051/0004-6361:20020519).
- de Vaucouleurs, G. (1948). Recherches sur les Nebuleuses Extragalactiques. *Annales d'Astrophysique*, 11, 247.
- Deane, R.P.; Paragi, Z.; Jarvis, M.J.; Coriat, M.; Bernardi, G.; Fender, R.P.; Frey, S.; Heywood, I.; Klöckner, H.R.; Grainge, K.; et al. (2014). A close-pair binary in a distant triple supermassive black hole system. *Nature*, 511, 57–60. doi:[10.1038/nature13454](https://doi.org/10.1038/nature13454).
- Dehnen, W. (1993). A Family of Potential-Density Pairs for Spherical Galaxies and Bulges. *Mon. Not. R. Astron. Soc.*, 265, 250. doi:[10.1093/mnras/265.1.250](https://doi.org/10.1093/mnras/265.1.250).
- Demorest, P.B.; Ferdman, R.D.; Gonzalez, M.E.; Nice, D.; Ransom, S.; Stairs, I.H.; Arzoumanian, Z.; Brazier, A.; Burke-Spolaor, S.; Chamberlin, S.J.; et al. (2013). Limits on the Stochastic Gravitational Wave Background from the North American Nanohertz Observatory for Gravitational Waves. *Astrophys. J.*, 762, 94. doi:[10.1088/0004-637X/762/2/94](https://doi.org/10.1088/0004-637X/762/2/94).
- Desvignes, G.; Caballero, R.N.; Lentati, L.; Verbiest, J.P.W.; Champion, D.J.; Stappers, B.W.; Janssen, G.H.; Lazarus, P.; Osłowski, S.; Babak, S.; et al. (2016). High-precision timing of 42 millisecond pulsars with the European Pulsar Timing Array. *Mon. Not. R. Astron. Soc.*, 458, 3341–3380. doi:[10.1093/mnras/stw483](https://doi.org/10.1093/mnras/stw483).
- Detweiler, S. (1979). Pulsar timing measurements and the search for gravitational waves. *Astrophys. J.*, 234, 1100–1104. doi:[10.1086/157593](https://doi.org/10.1086/157593).
- Devecchi, B.; Volonteri, M. (2009). Formation of the First Nuclear Clusters and Massive Black Holes at High Redshift. *Astrophys. J.*, 694, 302–313. doi:[10.1088/0004-637X/694/1/302](https://doi.org/10.1088/0004-637X/694/1/302).
- Di Matteo, T.; Springel, V.; Hernquist, L. (2005). Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature*, 433, 604–607. doi:[10.1038/nature03335](https://doi.org/10.1038/nature03335).

- Dijkstra, M.; Ferrara, A.; Mesinger, A. (2014). Feedback-regulated supermassive black hole seed formation. *Mon. Not. R. Astron. Soc.*, 442, 2036–2047. doi:[10.1093/mnras/stu1007](https://doi.org/10.1093/mnras/stu1007).
- Dolag, K.; Borgani, S.; Schindler, S.; Diaferio, A.; Bykov, A.M. (2008). Simulation Techniques for Cosmological Simulations. *Space Science Rev.*, 134, 229–268. doi:[10.1007/s11214-008-9316-5](https://doi.org/10.1007/s11214-008-9316-5).
- D’Orazio, D.J.; Haiman, Z.; Duffell, P.; MacFadyen, A.; Farris, B. (2016). A transition in circumbinary accretion discs at a binary mass ratio of 1:25. *Mon. Not. R. Astron. Soc.*, 459, 2379–2393. doi:[10.1093/mnras/stw792](https://doi.org/10.1093/mnras/stw792).
- D’Orazio, D.J.; Haiman, Z.; Schiminovich, D. (2015). Relativistic boost as the cause of periodicity in a massive black-hole binary candidate. *Nature*, 525, 351–353. doi:[10.1038/nature15262](https://doi.org/10.1038/nature15262).
- Dotti, M.; Colpi, M.; Haardt, F.; Mayer, L. (2007). Supermassive black hole binaries in gaseous and stellar circumnuclear discs: orbital dynamics and gas accretion. *Mon. Not. R. Astron. Soc.*, 379, 956–962. doi:[10.1111/j.1365-2966.2007.12010.x](https://doi.org/10.1111/j.1365-2966.2007.12010.x).
- Dotti, M.; Montuori, C.; Decarli, R.; Volonteri, M.; Colpi, M.; Haardt, F. (2009). SDSSJ092712.65+294344.0: a candidate massive black hole binary. *Mon. Not. R. Astron. Soc.*, 398, L73–L77. doi:[10.1111/j.1745-3933.2009.00714.x](https://doi.org/10.1111/j.1745-3933.2009.00714.x).
- Dotti, M.; Volonteri, M.; Perego, A.; Colpi, M.; Ruszkowski, M.; Haardt, F. (2010). Dual black holes in merger remnants - II. Spin evolution and gravitational recoil. *Mon. Not. R. Astron. Soc.*, 402, 682–690. doi:[10.1111/j.1365-2966.2009.15922.x](https://doi.org/10.1111/j.1365-2966.2009.15922.x).
- Ekers, R.D.; Fanti, R.; Lari, C.; Parma, P. (1978). NGC326 - A radio galaxy with a precessing beam. *Nature*, 276, 588–590. doi:[10.1038/276588a0](https://doi.org/10.1038/276588a0).
- Ellis, J.A. (2013). A Bayesian analysis pipeline for continuous GW sources in the PTA band. *Classical and Quantum Gravity*, 30, 22, 224004. doi:[10.1088/0264-9381/30/22/224004](https://doi.org/10.1088/0264-9381/30/22/224004).
- Ellis, J.A.; Siemens, X.; Creighton, J.D.E. (2012). Optimal Strategies for Continuous Gravitational Wave Detection in Pulsar Timing Arrays. *Astrophys. J.*, 756, 175. doi:[10.1088/0004-637X/756/2/175](https://doi.org/10.1088/0004-637X/756/2/175).
- Enoki, M.; Inoue, K.T.; Nagashima, M.; Sugiyama, N. (2004). Gravitational Waves from Supermassive Black Hole Coalescence in a Hierarchical Galaxy Formation Model. *Astrophys. J.*, 615, 19–28. doi:[10.1086/424475](https://doi.org/10.1086/424475).
- Enoki, M.; Nagashima, M. (2007). The Effect of Orbital Eccentricity on Gravitational Wave Background Radiation from Supermassive Black Hole Binaries. *Progress of Theoretical Physics*, 117, 241–256. doi:[10.1143/PTP.117.241](https://doi.org/10.1143/PTP.117.241).
- Eracleous, M.; Boroson, T.A.; Halpern, J.P.; Liu, J. (2012). A Large Systematic Search for Close Supermassive Binary and Rapidly Recoiling Black Holes. *Astrophys. J., Suppl. Ser.*, 201, 23. doi:[10.1088/0067-0049/201/2/23](https://doi.org/10.1088/0067-0049/201/2/23).
- Eracleous, M.; Halpern, J.P.; M. Gilbert, A.; Newman, J.A.; Filippenko, A.V. (1997). Rejection of the Binary Broad-Line Region Interpretation of Double-peaked Emission Lines in Three Active Galactic Nuclei. *Astrophys. J.*, 490, 216–226. doi:[10.1086/304859](https://doi.org/10.1086/304859).
- Escala, A.; Coppi, P.S.; Larson, R.B.; Mardones, D. (2005a). The Role of Gas in the Merging of Massive

- Black Holes in Galactic Nuclei. In A. Merloni; S. Nayakshin; R.A. Sunyaev (editors), *Growing Black Holes: Accretion in a Cosmological Context*. 352–353. doi:[10.1007/11403913\\_67](https://doi.org/10.1007/11403913_67).
- Escala, A.; Larson, R.B.; Coppi, P.S.; Mardones, D. (2005b). The Role of Gas in the Merging of Massive Black Holes in Galactic Nuclei. II. Black Hole Merging in a Nuclear Gas Disk. *Astrophys. J.*, 630, 152–166. doi:[10.1086/431747](https://doi.org/10.1086/431747).
- Estabrook, F.B.; Wahlquist, H.D. (1975). Response of Doppler spacecraft tracking to gravitational radiation. *General Relativity and Gravitation*, 6, 439–447. doi:[10.1007/BF00762449](https://doi.org/10.1007/BF00762449).
- Faber, S.M.; Tremaine, S.; Ajhar, E.A.; Byun, Y.I.; Dressler, A.; Gebhardt, K.; Grillmair, C.; Kormendy, J.; Lauer, T.R.; Richstone, D. (1997). The Centers of Early-Type Galaxies with HST. IV. Central Parameter Relations. *Astron. J.*, 114, 1771. doi:[10.1086/118606](https://doi.org/10.1086/118606).
- Fan, J.H.; Lin, R.G.; Xie, G.Z.; Zhang, L.; Mei, D.C.; Su, C.Y.; Peng, Z.M. (2002). Optical periodicity analysis for radio selected BL Lacertae objects (RBLs). *Astron. Astrophys.*, 381, 1–5. doi:[10.1051/0004-6361:20011356](https://doi.org/10.1051/0004-6361:20011356).
- Fan, X.; Strauss, M.A.; Richards, G.T.; Hennawi, J.F.; Becker, R.H.; White, R.L.; Diamond-Stanic, A.M.; Donley, J.L.; Jiang, L.; Kim, J.S.; et al. (2006). A Survey of  $z > 5.7$  Quasars in the Sloan Digital Sky Survey. IV. Discovery of Seven Additional Quasars. *Astron. J.*, 131, 1203–1209. doi:[10.1086/500296](https://doi.org/10.1086/500296).
- Farris, B.D.; Duffell, P.; MacFadyen, A.I.; Haiman, Z. (2014). Binary Black Hole Accretion from a Circumbinary Disk: Gas Dynamics inside the Central Cavity. *Astrophys. J.*, 783, 134. doi:[10.1088/0004-637X/783/2/134](https://doi.org/10.1088/0004-637X/783/2/134).
- Ferrarese, L.; Ford, H. (2005). Supermassive Black Holes in Galactic Nuclei: Past, Present and Future Research. *Space Science Rev.*, 116, 523–624. doi:[10.1007/s11214-005-3947-6](https://doi.org/10.1007/s11214-005-3947-6).
- Ferrarese, L.; Merritt, D. (2000). A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies. *Astrophys. J. Lett.*, 539, L9–L12. doi:[10.1086/312838](https://doi.org/10.1086/312838).
- Fiedler, R.A.; Mouschovias, T.C. (1992). Ambipolar diffusion and star formation: Formation and contraction of axisymmetric cloud cores. I - Formulation of the problem and method of solution. *Astrophys. J.*, 391, 199–219. doi:[10.1086/171336](https://doi.org/10.1086/171336).
- Finn, L.S.; Larson, S.L.; Romano, J.D. (2009). Detecting a stochastic gravitational-wave background: The overlap reduction function. *Phys. Rev. D*, 79, 6, 062003. doi:[10.1103/PhysRevD.79.062003](https://doi.org/10.1103/PhysRevD.79.062003).
- Foreman, G.; Volonteri, M.; Dotti, M. (2009). Double Quasars: Probes of Black Hole Scaling Relationships and Merger Scenarios. *Astrophys. J.*, 693, 1554–1562. doi:[10.1088/0004-637X/693/2/1554](https://doi.org/10.1088/0004-637X/693/2/1554).
- Foster, R.S.; Backer, D.C. (1990). Constructing a pulsar timing array. *Astrophys. J.*, 361, 300–308. doi:[10.1086/169195](https://doi.org/10.1086/169195).
- Frank, J.; King, A.; Raine, D.J. (2002). *Accretion Power in Astrophysics: Third Edition*.
- Frank, J.; Rees, M.J. (1976). Effects of massive central black holes on dense stellar systems. *Mon. Not. R. Astron. Soc.*, 176, 633–647.
- Fu, H.; Myers, A.D.; Djorgovski, S.G.; Yan, L. (2011a). Mergers in Double-peaked [O III] Active Galactic

Nuclei. *Astrophys. J.*, 733, 103. doi:[10.1088/0004-637X/733/2/103](https://doi.org/10.1088/0004-637X/733/2/103).

- Fu, H.; Zhang, Z.Y.; Assef, R.J.; Stockton, A.; Myers, A.D.; Yan, L.; Djorgovski, S.G.; Wrobel, J.M.; Riechers, D.A. (2011b). A Kiloparsec-scale Binary Active Galactic Nucleus Confirmed by the Expanded Very Large Array. *Astrophys. J. Lett.*, 740, L44. doi:[10.1088/2041-8205/740/2/L44](https://doi.org/10.1088/2041-8205/740/2/L44).
- Gaskell, C.M. (1983). Quasars as supermassive binaries. In J.P. Swings (editor), *Liege International Astrophysical Colloquia*, volume 24 of *Liege International Astrophysical Colloquia*. 473–477.
- Gebhardt, K.; Bender, R.; Bower, G.; Dressler, A.; Faber, S.M.; Filippenko, A.V.; Green, R.; Grillmair, C.; Ho, L.C.; Kormendy, J.; et al. (2000). A Relationship between Nuclear Black Hole Mass and Galaxy Velocity Dispersion. *Astrophys. J. Lett.*, 539, L13–L16. doi:[10.1086/312840](https://doi.org/10.1086/312840).
- Genel, S.; Vogelsberger, M.; Springel, V.; Sijacki, D.; Nelson, D.; Snyder, G.; Rodriguez-Gomez, V.; Torrey, P.; Hernquist, L. (2014). Introducing the Illustris project: the evolution of galaxy populations across cosmic time. *Mon. Not. R. Astron. Soc.*, 445, 175–200. doi:[10.1093/mnras/stu1654](https://doi.org/10.1093/mnras/stu1654).
- Gingold, R.A.; Monaghan, J.J. (1977). Smoothed particle hydrodynamics - Theory and application to non-spherical stars. *Mon. Not. R. Astron. Soc.*, 181, 375–389. doi:[10.1093/mnras/181.3.375](https://doi.org/10.1093/mnras/181.3.375).
- Gnedin, N.Y. (1995). Softened Lagrangian hydrodynamics for cosmology. *Astrophys. J., Suppl. Ser.*, 97, 231–257. doi:[10.1086/192141](https://doi.org/10.1086/192141).
- Goicovic, F.G.; Sesana, A.; Cuadra, J.; Stasyszyn, F. (2017). Infalling clouds on to supermassive black hole binaries - II. Binary evolution and the final parsec problem. *Mon. Not. R. Astron. Soc.*, 472, 514–531. doi:[10.1093/mnras/stx1996](https://doi.org/10.1093/mnras/stx1996).
- Gould, A.; Rix, H.W. (2000). Binary Black Hole Mergers from Planet-like Migrations. *Astrophys. J. Lett.*, 532, L29–L32. doi:[10.1086/312562](https://doi.org/10.1086/312562).
- Goulding, A.D.; Greene, J.E.; Bezanson, R.; Greco, J.; Johnson, S.; Leauthaud, A.; Matsuoka, Y.; Medezinski, E.; Price-Whelan, A.M. (2018). Galaxy interactions trigger rapid black hole growth: An unprecedented view from the Hyper Suprime-Cam survey. *PASJ*, 70, S37. doi:[10.1093/pasj/psx135](https://doi.org/10.1093/pasj/psx135).
- Graham, M.J.; Djorgovski, S.G.; Stern, D.; Drake, A.J.; Mahabal, A.A.; Donalek, C.; Glikman, E.; Larson, S.; Christensen, E. (2015a). A systematic search for close supermassive black hole binaries in the Catalina Real-time Transient Survey. *Mon. Not. R. Astron. Soc.*, 453, 1562–1576. doi:[10.1093/mnras/stv1726](https://doi.org/10.1093/mnras/stv1726).
- Graham, M.J.; Djorgovski, S.G.; Stern, D.; Glikman, E.; Drake, A.J.; Mahabal, A.A.; Donalek, C.; Larson, S.; Christensen, E. (2015b). A possible close supermassive black-hole binary in a quasar with optical periodicity. *Nature*, 518, 74–76. doi:[10.1038/nature14143](https://doi.org/10.1038/nature14143).
- Guo, Q.; White, S.; Boylan-Kolchin, M.; De Lucia, G.; Kauffmann, G.; Lemson, G.; Li, C.; Springel, V.; Weinmann, S. (2011). From dwarf spheroidals to cD galaxies: simulating the galaxy population in a  $\Lambda$ CDM cosmology. *Mon. Not. R. Astron. Soc.*, 413, 101–131. doi:[10.1111/j.1365-2966.2010.18114.x](https://doi.org/10.1111/j.1365-2966.2010.18114.x).
- Haiman, Z. (2013). The Formation of the First Massive Black Holes. In T. Wiklind; B. Mobasher; V. Bromm (editors), *The First Galaxies*, volume 396 of *Astrophysics and Space Science Library*. 293. doi:[10.1007/978-3-642-32362-1\\_6](https://doi.org/10.1007/978-3-642-32362-1_6).



- Haiman, Z.; Kocsis, B.; Menou, K. (2009). The Population of Viscosity- and Gravitational Wave-driven Supermassive Black Hole Binaries Among Luminous Active Galactic Nuclei. *Astrophys. J.*, 700, 1952–1969. doi:[10.1088/0004-637X/700/2/1952](https://doi.org/10.1088/0004-637X/700/2/1952).
- Halpern, J.P.; Filippenko, A.V. (1988). A test of the massive binary black hole hypothesis - ARP 102B. *Nature*, 331, 46–48. doi:[10.1038/331046a0](https://doi.org/10.1038/331046a0).
- Hashimoto, Y.; Funato, Y.; Makino, J. (2003). To Circularize or Not To Circularize?-Orbital Evolution of Satellite Galaxies. *Astrophys. J.*, 582, 196–201. doi:[10.1086/344260](https://doi.org/10.1086/344260).
- Hayasaki, K. (2009). A New Mechanism for Massive Binary Black-Hole Evolution. *PASJ*, 61, 65–74. doi:[10.1093/pasj/61.1.65](https://doi.org/10.1093/pasj/61.1.65).
- Heggie, D.C. (1975). Binary evolution in stellar dynamics. *Mon. Not. R. Astron. Soc.*, 173, 729–787. doi:[10.1093/mnras/173.3.729](https://doi.org/10.1093/mnras/173.3.729).
- Hellings, R.W.; Downs, G.S. (1983). Upper limits on the isotropic gravitational radiation background from pulsar timing analysis. *Astrophys. J. Lett.*, 265, L39–L42. doi:[10.1086/183954](https://doi.org/10.1086/183954).
- Hernquist, L. (1990). An analytical model for spherical galaxies and bulges. *Astrophys. J.*, 356, 359–364. doi:[10.1086/168845](https://doi.org/10.1086/168845).
- Hernquist, L. (1993). Some cautionary remarks about smoothed particle hydrodynamics. *Astrophys. J.*, 404, 717–722. doi:[10.1086/172325](https://doi.org/10.1086/172325).
- Hernquist, L.; Katz, N. (1989). TREESPH - A unification of SPH with the hierarchical tree method. *Astrophys. J., Suppl. Ser.*, 70, 419–446. doi:[10.1086/191344](https://doi.org/10.1086/191344).
- Hills, J.G. (1975). Encounters between binary and single stars and their effect on the dynamical evolution of stellar systems. *Astron. J.*, 80, 809–825. doi:[10.1086/111815](https://doi.org/10.1086/111815).
- Hobbs, G.; Archibald, A.; Arzoumanian, Z.; Backer, D.; Bailes, M.; Bhat, N.D.R.; Burgay, M.; Burke-Spolaor, S.; Champion, D.; Cognard, I.; et al. (2010). The International Pulsar Timing Array project: using pulsars as a gravitational wave detector. *Classical and Quantum Gravity*, 27, 8, 084013. doi:[10.1088/0264-9381/27/8/084013](https://doi.org/10.1088/0264-9381/27/8/084013).
- Hockney, R.W.; Goel, S.P.; Eastwood, J.W. (1974). Quiet High-Resolution Computer Models of a Plasma. *Journal of Computational Physics*, 14, 148–158. doi:[10.1016/0021-9991\(74\)90010-2](https://doi.org/10.1016/0021-9991(74)90010-2).
- Hoffman, L.; Loeb, A. (2007). Dynamics of triple black hole systems in hierarchically merging massive galaxies. *Mon. Not. R. Astron. Soc.*, 377, 957–976. doi:[10.1111/j.1365-2966.2007.11694.x](https://doi.org/10.1111/j.1365-2966.2007.11694.x).
- Hogg, D.W. (1999). Distance measures in cosmology. *ArXiv Astrophysics e-prints*.
- Hopkins, P.F.; Croton, D.; Bundy, K.; Khochfar, S.; van den Bosch, F.; Somerville, R.S.; Wetzell, A.; Keres, D.; Hernquist, L.; Stewart, K.; et al. (2010). Mergers in  $\Lambda$ CDM: Uncertainties in Theoretical Predictions and Interpretations of the Merger Rate. *Astrophys. J.*, 724, 915–945. doi:[10.1088/0004-637X/724/2/915](https://doi.org/10.1088/0004-637X/724/2/915).
- Hopkins, P.F.; Richards, G.T.; Hernquist, L. (2007). An Observational Determination of the Bolometric Quasar Luminosity Function. *Astrophys. J.*, 654, 731–753. doi:[10.1086/509629](https://doi.org/10.1086/509629).

- Huerta, E.A.; McWilliams, S.T.; Gair, J.R.; Taylor, S.R. (2015). Detection of eccentric supermassive black hole binaries with pulsar timing arrays: Signal-to-noise ratio calculations. *Phys. Rev. D*, 92, 6, 063010. doi:[10.1103/PhysRevD.92.063010](https://doi.org/10.1103/PhysRevD.92.063010).
- Hunter, J.D. (2007). Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9, 3, 90–95.
- Hut, P. (1983). Binaries as a heat source in stellar dynamics - Release of binding energy. *Astrophys. J. Lett.*, 272, L29–L33. doi:[10.1086/184111](https://doi.org/10.1086/184111).
- Illingworth, G. (1977). Rotation in 13 elliptical galaxies. *Astrophys. J. Lett.*, 218, L43–L47. doi:[10.1086/182572](https://doi.org/10.1086/182572).
- Ivanov, P.B.; Polnarev, A.G.; Saha, P. (2005a). The tidal disruption rate in dense galactic cusps containing a supermassive binary black hole. *Mon. Not. R. Astron. Soc.*, 358, 1361–1378. doi:[10.1111/j.1365-2966.2005.08843.x](https://doi.org/10.1111/j.1365-2966.2005.08843.x).
- Ivanov, P.B.; Polnarev, A.G.; Saha, P. (2005b). The tidal disruption rate in dense galactic cusps containing a supermassive binary black hole. *Mon. Not. R. Astron. Soc.*, 358, 1361–1378. doi:[10.1111/j.1365-2966.2005.08843.x](https://doi.org/10.1111/j.1365-2966.2005.08843.x).
- Ivezic, Z.; Tyson, J.A.; Abel, B.; Acosta, E.; Allsman, R.; AlSayyad, Y.; Anderson, S.F.; Andrew, J.; Angel, R.; Angeli, G.; et al. (2008). LSST: from Science Drivers to Reference Design and Anticipated Data Products. *ArXiv e-prints*.
- Jaffe, A.H.; Backer, D.C. (2003). Gravitational Waves Probe the Coalescence Rate of Massive Black Hole Binaries. *The Astrophysical Journal*, 583, 2, 616–631. doi:[10.1086/345443](https://doi.org/10.1086/345443).
- Janet, F.A.; Hobbs, G.B.; Lee, K.J.; Manchester, R.N. (2005). Detecting the Stochastic Gravitational Wave Background Using Pulsar Timing. *Astrophys. J. Lett.*, 625, L123–L126. doi:[10.1086/431220](https://doi.org/10.1086/431220).
- Jiang, Y.F.; Stone, J.M.; Davis, S.W. (2014). A Global Three-dimensional Radiation Magneto-hydrodynamic Simulation of Super-Eddington Accretion Disks. *Astrophys. J.*, 796, 106. doi:[10.1088/0004-637X/796/2/106](https://doi.org/10.1088/0004-637X/796/2/106).
- Jones, E.; Oliphant, T.; Peterson, P.; et al. (2001). SciPy: Open source scientific tools for Python. [Online; accessed 2016-01-28].
- Kelley, L.Z.; Blecha, L.; Hernquist, L. (2017). Massive black hole binary mergers in dynamical galactic environments. *Mon. Not. R. Astron. Soc.*, 464, 3131–3157. doi:[10.1093/mnras/stw2452](https://doi.org/10.1093/mnras/stw2452).
- Kelley, L.Z.; Blecha, L.; Hernquist, L.; Sesana, A.; Taylor, S.R. (2017). The gravitational wave background from massive black hole binaries in Illustris: spectral features and time to detection with pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 471, 4508–4526. doi:[10.1093/mnras/stx1638](https://doi.org/10.1093/mnras/stx1638).
- Kelley, L.Z.; Blecha, L.; Hernquist, L.; Sesana, A.; Taylor, S.R. (2018a). Single Sources in the Low-Frequency Gravitational Wave Sky: properties and time to detection by pulsar timing arrays. *Mon. Not. R. Astron. Soc.*. doi:[10.1093/mnras/sty689](https://doi.org/10.1093/mnras/sty689).
- Kelley, L.Z.; Haiman, Z.; Sesana, A.; Hernquist, L. (2018b). Massive BH Binaries as Periodically-Variable AGN. *in prep*.

- Kepler, J. (1619). *Harmonices mundi libri V*.
- Khan, F.M.; Fiacconi, D.; Mayer, L.; Berczik, P.; Just, A. (2016). Swift coalescence of supermassive black holes in cosmological mergers of massive galaxies. *ArXiv e-prints*.
- Khan, F.M.; Holley-Bockelmann, K.; Berczik, P.; Just, A. (2013). Supermassive Black Hole Binary Evolution in Axisymmetric Galaxies: The Final Parsec Problem is Not a Problem. *Astrophys. J.*, 773, 100. doi:[10.1088/0004-637X/773/2/100](https://doi.org/10.1088/0004-637X/773/2/100).
- Khan, F.M.; Just, A.; Merritt, D. (2011). Efficient Merger of Binary Supermassive Black Holes in Merging Galaxies. *Astrophys. J.*, 732, 89. doi:[10.1088/0004-637X/732/2/89](https://doi.org/10.1088/0004-637X/732/2/89).
- Kocsis, B.; Sesana, A. (2011). Gas-driven massive black hole binaries: signatures in the nHz gravitational wave background. *Mon. Not. R. Astron. Soc.*, 411, 1467–1479. doi:[10.1111/j.1365-2966.2010.17782.x](https://doi.org/10.1111/j.1365-2966.2010.17782.x).
- Komossa, S. (2006). Observational evidence for binary black holes and active double nuclei. *Mem. Societa Astronomica Italiana*, 77, 733.
- Komossa, S.; Burwitz, V.; Hasinger, G.; Predehl, P.; Kaastra, J.S.; Ikebe, Y. (2003). Discovery of a Binary Active Galactic Nucleus in the Ultraluminous Infrared Galaxy NGC 6240 Using Chandra. *Astrophys. J. Lett.*, 582, L15–L19. doi:[10.1086/346145](https://doi.org/10.1086/346145).
- Kormendy, J.; Richstone, D. (1995). Inward Bound—The Search For Supermassive Black Holes In Galactic Nuclei. *Annu. Rev. Astron. Astrophys.*, 33, 581. doi:[10.1146/annurev.aa.33.090195.003053](https://doi.org/10.1146/annurev.aa.33.090195.003053).
- Koss, M.; Mushotzky, R.; Treister, E.; Veilleux, S.; Vasudevan, R.; Trippe, M. (2012). Understanding Dual Active Galactic Nucleus Activation in the nearby Universe. *Astrophys. J. Lett.*, 746, L22. doi:[10.1088/2041-8205/746/2/L22](https://doi.org/10.1088/2041-8205/746/2/L22).
- Kozai, Y. (1962). Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.*, 67, 591. doi:[10.1086/108790](https://doi.org/10.1086/108790).
- Kramer, M.; Champion, D.J. (2013). The European Pulsar Timing Array and the Large European Array for Pulsars. *Classical and Quantum Gravity*, 30, 22, 224009. doi:[10.1088/0264-9381/30/22/224009](https://doi.org/10.1088/0264-9381/30/22/224009).
- Kulier, A.; Ostriker, J.P.; Natarajan, P.; Lackner, C.N.; Cen, R. (2015). Understanding Black Hole Mass Assembly via Accretion and Mergers at Late Times in Cosmological Simulations. *Astrophys. J.*, 799, 178. doi:[10.1088/0004-637X/799/2/178](https://doi.org/10.1088/0004-637X/799/2/178).
- Lacey, C.; Cole, S. (1993). Merger rates in hierarchical models of galaxy formation. *Mon. Not. R. Astron. Soc.*, 262, 627–649. doi:[10.1093/mnras/262.3.627](https://doi.org/10.1093/mnras/262.3.627).
- Lauer, T.R.; Gebhardt, K.; Faber, S.M.; Richstone, D.; Tremaine, S.; Kormendy, J.; Aller, M.C.; Bender, R.; Dressler, A.; Filippenko, A.V.; et al. (2007). The Centers of Early-Type Galaxies with Hubble Space Telescope. VI. Bimodal Central Surface Brightness Profiles. *Astrophys. J.*, 664, 226–256. doi:[10.1086/519229](https://doi.org/10.1086/519229).
- Lauer, T.R.; Gebhardt, K.; Richstone, D.; Tremaine, S.; Bender, R.; Bower, G.; Dressler, A.; Faber, S.M.; Filippenko, A.V.; Green, R.; et al. (2002). Galaxies with a Central Minimum in Stellar Luminosity Density. *Astron. J.*, 124, 1975–1987. doi:[10.1086/342932](https://doi.org/10.1086/342932).

- Leach, R. (1981). On the triaxiality of elliptical galaxies. *Astrophys. J.*, 248, 485–498. doi:[10.1086/159173](https://doi.org/10.1086/159173).
- Lee, K.J.; Wex, N.; Kramer, M.; Stappers, B.W.; Bassa, C.G.; Janssen, G.H.; Karuppusamy, R.; Smits, R. (2011). Gravitational wave astronomy of single sources with a pulsar timing array. *Mon. Not. R. Astron. Soc.*, 414, 3251–3264. doi:[10.1111/j.1365-2966.2011.18622.x](https://doi.org/10.1111/j.1365-2966.2011.18622.x).
- Lentati, L.; Shannon, R.M.; Coles, W.A.; Verbiest, J.P.W.; van Haasteren, R.; Ellis, J.A.; Caballero, R.N.; Manchester, R.N.; Arzoumanian, Z.; Babak, S.; et al. (2016). From spin noise to systematics: stochastic processes in the first International Pulsar Timing Array data release. *Mon. Not. R. Astron. Soc.*, 458, 2161–2187. doi:[10.1093/mnras/stw395](https://doi.org/10.1093/mnras/stw395).
- Lentati, L.; Taylor, S.R.; Mingarelli, C.M.F.; Sesana, A.; Sanidas, S.A.; Vecchio, A.; Caballero, R.N.; Lee, K.J.; van Haasteren, R.; Babak, S.; et al. (2015). European Pulsar Timing Array limits on an isotropic stochastic gravitational-wave background. *Mon. Not. R. Astron. Soc.*, 453, 2576–2598. doi:[10.1093/mnras/stv1538](https://doi.org/10.1093/mnras/stv1538).
- Li, G.; Naoz, S.; Kocsis, B.; Loeb, A. (2015). Implications of the eccentric Kozai-Lidov mechanism for stars surrounding supermassive black hole binaries. *Mon. Not. R. Astron. Soc.*, 451, 1341–1349. doi:[10.1093/mnras/stv1031](https://doi.org/10.1093/mnras/stv1031).
- Li, S.; Liu, F.K.; Berczik, P.; Spurzem, R. (2017). Boosted Tidal Disruption by Massive Black Hole Binaries During Galaxy Mergers from the View of N-Body Simulation. *Astrophys. J.*, 834, 195. doi:[10.3847/1538-4357/834/2/195](https://doi.org/10.3847/1538-4357/834/2/195).
- Lidov, M.L. (1962). The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.*, 9, 719–759. doi:[10.1016/0032-0633\(62\)90129-0](https://doi.org/10.1016/0032-0633(62)90129-0).
- Lightman, A.P.; Shapiro, S.L. (1977). The distribution and consumption rate of stars around a massive, collapsed object. *Astrophys. J.*, 211, 244–262. doi:[10.1086/154925](https://doi.org/10.1086/154925).
- LIGO (2016). GW150914: The Advanced LIGO Detectors in the Era of First Discoveries. *ArXiv e-prints*.
- Liu, T.; Gezari, S.; Burgett, W.; Chambers, K.; Draper, P.; Hodapp, K.; Huber, M.; Kudritzki, R.P.; Magnier, E.; Metcalfe, N.; et al. (2016). A Systematic Search for Periodically Varying Quasars in Pan-STARRS1: An Extended Baseline Test in Medium Deep Survey Field MD09. *Astrophys. J.*, 833, 6. doi:[10.3847/0004-637X/833/1/6](https://doi.org/10.3847/0004-637X/833/1/6).
- Liu, T.; Gezari, S.; Miller, M.C. (2018). Did ASAS-SN Kill the Supermassive Black Hole Binary Candidate PG1302-102? *ArXiv e-prints*.
- Liu, X.; Greene, J.E.; Shen, Y.; Strauss, M.A. (2010). Discovery of Four kpc-scale Binary Active Galactic Nuclei. *Astrophys. J. Lett.*, 715, L30–L34. doi:[10.1088/2041-8205/715/1/L30](https://doi.org/10.1088/2041-8205/715/1/L30).
- Liu, X.; Shen, Y.; Bian, F.; Loeb, A.; Tremaine, S. (2014). Constraining Sub-parsec Binary Supermassive Black Holes in Quasars with Multi-epoch Spectroscopy. II. The Population with Kinematically Offset Broad Balmer Emission Lines. *Astrophys. J.*, 789, 140. doi:[10.1088/0004-637X/789/2/140](https://doi.org/10.1088/0004-637X/789/2/140).
- Lotz, J.M.; Jonsson, P.; Cox, T.J.; Croton, D.; Primack, J.R.; Somerville, R.S.; Stewart, K. (2011). The Major and Minor Galaxy Merger Rates at  $z < 1.5$ . *Astrophys. J.*, 742, 103. doi:[10.1088/0004-637X/742/2/103](https://doi.org/10.1088/0004-637X/742/2/103).

- Lucy, L.B. (1977). A numerical approach to the testing of the fission hypothesis. *Astron. J.*, 82, 1013–1024. doi:[10.1086/112164](https://doi.org/10.1086/112164).
- Lynden-Bell, D. (1969). Galactic Nuclei as Collapsed Old Quasars. *Nature*, 223, 690–694. doi:[10.1038/223690a0](https://doi.org/10.1038/223690a0).
- Madau, P.; Haardt, F.; Dotti, M. (2014). Super-critical Growth of Massive Black Holes from Stellar-mass Seeds. *Astrophys. J.*, 784. doi:[10.1088/2041-8205/784/2/L38](https://doi.org/10.1088/2041-8205/784/2/L38).
- Madau, P.; Rees, M.J. (2001). Massive Black Holes as Population III Remnants. *Astrophys. J. Lett.*, 551, L27–L30. doi:[10.1086/319848](https://doi.org/10.1086/319848).
- Magorrian, J.; Tremaine, S. (1999). Rates of tidal disruption of stars by massive central black holes. *Mon. Not. R. Astron. Soc.*, 309, 447–460. doi:[10.1046/j.1365-8711.1999.02853.x](https://doi.org/10.1046/j.1365-8711.1999.02853.x).
- Magorrian, J.; Tremaine, S.; Richstone, D.; Bender, R.; Bower, G.; Dressler, A.; Faber, S.M.; Gebhardt, K.; Green, R.; Grillmair, C.; et al. (1998). The Demography of Massive Dark Objects in Galaxy Centers. *Astron. J.*, 115, 2285–2305. doi:[10.1086/300353](https://doi.org/10.1086/300353).
- Mahadevan, R. (1997). Scaling Laws for Advection-dominated Flows: Applications to Low-Luminosity Galactic Nuclei. *Astrophys. J.*, 477, 585–601. doi:[10.1086/303727](https://doi.org/10.1086/303727).
- Manchester, R.N.; Hobbs, G.; Bailes, M.; Coles, W.A.; van Straten, W.; Keith, M.J.; Shannon, R.M.; Bhat, N.D.R.; Brown, A.; Burke-Spolaor, S.G.; et al. (2013). The Parkes Pulsar Timing Array Project. *PASA*, 30, e017. doi:[10.1017/pasa.2012.017](https://doi.org/10.1017/pasa.2012.017).
- Marconi, A.; Hunt, L.K. (2003). The Relation between Black Hole Mass, Bulge Mass, and Near-Infrared Luminosity. *Astrophys. J. Lett.*, 589, L21–L24. doi:[10.1086/375804](https://doi.org/10.1086/375804).
- McConnell, N.J.; Ma, C.P. (2013). Revisiting the Scaling Relations of Black Hole Masses and Host Galaxy Properties. *Astrophys. J.*, 764, 184. doi:[10.1088/0004-637X/764/2/184](https://doi.org/10.1088/0004-637X/764/2/184).
- McLaughlin, M.A. (2013). The North American Nanohertz Observatory for Gravitational Waves. *Classical and Quantum Gravity*, 30, 22, 224008. doi:[10.1088/0264-9381/30/22/224008](https://doi.org/10.1088/0264-9381/30/22/224008).
- McWilliams, S.T.; Ostriker, J.P.; Pretorius, F. (2014). Gravitational Waves and Stalled Satellites from Massive Galaxy Mergers at  $z \leq 1$ . *Astrophys. J.*, 789, 156. doi:[10.1088/0004-637X/789/2/156](https://doi.org/10.1088/0004-637X/789/2/156).
- Mendel, J.T.; Simard, L.; Palmer, M.; Ellison, S.L.; Patton, D.R. (2014). A Catalog of Bulge, Disk, and Total Stellar Mass Estimates for the Sloan Digital Sky Survey. *Astrophys. J., Suppl. Ser.*, 210, 3. doi:[10.1088/0067-0049/210/1/3](https://doi.org/10.1088/0067-0049/210/1/3).
- Merritt, D. (2013). Loss-cone dynamics. *Classical and Quantum Gravity*, 30, 24, 244005.
- Merritt, D.; Ekers, R.D. (2002). Tracing Black Hole Mergers Through Radio Lobe Morphology. *Science*, 297, 1310–1313. doi:[10.1126/science.1074688](https://doi.org/10.1126/science.1074688).
- Merritt, D.; Mikkola, S.; Szell, A. (2007). Long-Term Evolution of Massive Black Hole Binaries. III. Binary Evolution in Collisional Nuclei. *Astrophys. J.*, 671, 53–72. doi:[10.1086/522691](https://doi.org/10.1086/522691).
- Merritt, D.; Milosavljević, M. (2005). Massive Black Hole Binary Evolution. *Living Reviews in Relativity*,

8. doi:[10.12942/lrr-2005-8](https://doi.org/10.12942/lrr-2005-8).

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

Miller, M.C.; Krolik, J.H. (2013). Alignment of Supermassive Black Hole Binary Orbits and Spins. *Astrophys. J.*, 774, 43. doi:[10.1088/0004-637X/774/1/43](https://doi.org/10.1088/0004-637X/774/1/43).

Milosavljević, M.; Merritt, D. (2003). Long-Term Evolution of Massive Black Hole Binaries. *Astrophys. J.*, 596, 860–878. doi:[10.1086/378086](https://doi.org/10.1086/378086).

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

Monaghan, J.J. (1992). Smoothed particle hydrodynamics. *Annu. Rev. Astron. Astrophys.*, 30, 543–574. doi:[10.1146/annurev.aa.30.090192.002551](https://doi.org/10.1146/annurev.aa.30.090192.002551).

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

Muñoz, D.J.; Lai, D. (2016). Pulsed Accretion onto Eccentric and Circular Binaries. *Astrophys. J.*, 827, 43. doi:[10.3847/0004-637X/827/1/43](https://doi.org/10.3847/0004-637X/827/1/43).

Narayan, R.; Yi, I. (1995). Advection-dominated Accretion: Underfed Black Holes and Neutron Stars. *Astrophys. J.*, 452, 710. doi:[10.1086/176343](https://doi.org/10.1086/176343).

Nelson, D.; Pillepich, A.; Genel, S.; Vogelsberger, M.; Springel, V.; Torrey, P.; Rodriguez-Gomez, V.; Sijacki, D.; Snyder, G.F.; Griffen, B.; et al. (2015). The illustris simulation: Public data release. *Astronomy and Computing*, 13, 12–37. doi:[10.1016/j.ascom.2015.09.003](https://doi.org/10.1016/j.ascom.2015.09.003).

Noble, S.C.; Krolik, J.H.; Schnittman, J.D.; Hawley, J.F. (2011). Radiative Efficiency and Thermal Spectrum of Accretion onto Schwarzschild Black Holes. *Astrophys. J.*, 743, 115. doi:[10.1088/0004-637X/743/2/115](https://doi.org/10.1088/0004-637X/743/2/115).

Peebles, P.J.E. (1980). *The large-scale structure of the universe*.

Pen, U.L. (1998). A High-Resolution Adaptive Moving Mesh Hydrodynamic Algorithm. *Astrophys. J., Suppl. Ser.*, 115, 19–34. doi:[10.1086/313074](https://doi.org/10.1086/313074).

Peters, P.C. (1964). Gravitational Radiation and the Motion of Two Point Masses. *Physical Review*, 136, 1224–1232. doi:[10.1103/PhysRev.136.B1224](https://doi.org/10.1103/PhysRev.136.B1224).

Peters, P.C.; Mathews, J. (1963). Gravitational Radiation from Point Masses in a Keplerian Orbit. *Physical Review*, 131, 435–440. doi:[10.1103/PhysRev.131.435](https://doi.org/10.1103/PhysRev.131.435).

Petiteau, A.; Babak, S.; Sesana, A.; de Araújo, M. (2013). Resolving multiple supermassive black hole binaries with pulsar timing arrays. II. Genetic algorithm implementation. *Phys. Rev. D*, 87, 6, 064036. doi:[10.1103/PhysRevD.87.064036](https://doi.org/10.1103/PhysRevD.87.064036).

Phinney, E.S. (2001). A Practical Theorem on Gravitational Wave Backgrounds. *ArXiv Astrophysics*

*e-prints.*

- Press, W.H.; Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *Astrophys. J.*, 187, 425–438. doi:[10.1086/152650](https://doi.org/10.1086/152650).
- Preto, M.; Berentzen, I.; Berczik, P.; Spurzem, R. (2011). Fast Coalescence of Massive Black Hole Binaries from Mergers of Galactic Nuclei: Implications for Low-frequency Gravitational-wave Astrophysics. *Astrophys. J. Lett.*, 732, L26. doi:[10.1088/2041-8205/732/2/L26](https://doi.org/10.1088/2041-8205/732/2/L26).
- Pshirkov, M.S.; Baskaran, D.; Postnov, K.A. (2010). Observing gravitational wave bursts in pulsar timing measurements. *Mon. Not. R. Astron. Soc.*, 402, 417–423. doi:[10.1111/j.1365-2966.2009.15887.x](https://doi.org/10.1111/j.1365-2966.2009.15887.x).
- Prez, F.; Granger, B. (2007). Ipython: A system for interactive scientific computing. *Computing in Science Engineering*, 9, 3, 21–29. doi:[10.1109/MCSE.2007.53](https://doi.org/10.1109/MCSE.2007.53).
- Quinlan, G.D. (1996). The dynamical evolution of massive black hole binaries I. Hardening in a fixed stellar background. *Nature Astron.*, 1, 35–56. doi:[10.1016/S1384-1076\(96\)00003-6](https://doi.org/10.1016/S1384-1076(96)00003-6).
- Quinlan, G.D.; Hernquist, L. (1997). The dynamical evolution of massive black hole binaries — II. Self-consistent N-body integrations. *Nature Astron.*, 2, 533–554. doi:[10.1016/S1384-1076\(97\)00039-0](https://doi.org/10.1016/S1384-1076(97)00039-0).
- Rajagopal, M.; Romani, R.W. (1995). Ultra-Low-Frequency Gravitational Radiation from Massive Black Hole Binaries. *Astrophys. J.*, 446, 543. doi:[10.1086/175813](https://doi.org/10.1086/175813).
- Rantala, A.; Pihajoki, P.; Johansson, P.H.; Naab, T.; Lahén, N.; Sawala, T. (2016). Post-Newtonian dynamical modeling of supermassive black holes in galactic-scale simulations. *ArXiv e-prints*.
- Rasskazov, A.; Merritt, D. (2016). Evolution Of Massive Black Hole Binaries In Rotating Stellar Nuclei: Implications For Gravitational Wave Detection. *ArXiv e-prints*.
- Ravi, V.; Wyithe, J.S.B.; Hobbs, G.; Shannon, R.M.; Manchester, R.N.; Yardley, D.R.B.; Keith, M.J. (2012). Does a “Stochastic” Background of Gravitational Waves Exist in the Pulsar Timing Band? *Astrophys. J.*, 761, 84. doi:[10.1088/0004-637X/761/2/84](https://doi.org/10.1088/0004-637X/761/2/84).
- Ravi, V.; Wyithe, J.S.B.; Shannon, R.M.; Hobbs, G.; Manchester, R.N. (2014). Binary supermassive black hole environments diminish the gravitational wave signal in the pulsar timing band. *Mon. Not. R. Astron. Soc.*, 442, 56–68. doi:[10.1093/mnras/stu779](https://doi.org/10.1093/mnras/stu779).
- Reardon, D.J.; Hobbs, G.; Coles, W.; Levin, Y.; Keith, M.J.; Bailes, M.; Bhat, N.D.R.; Burke-Spolaor, S.; Dai, S.; Kerr, M.; et al. (2016). Timing analysis for 20 millisecond pulsars in the Parkes Pulsar Timing Array. *Mon. Not. R. Astron. Soc.*, 455, 1751–1769. doi:[10.1093/mnras/stv2395](https://doi.org/10.1093/mnras/stv2395).
- Rodriguez, C.; Taylor, G.B.; Zavala, R.T.; Peck, A.B.; Pollack, L.K.; Romani, R.W. (2006). A Compact Supermassive Binary Black Hole System. *Astrophys. J.*, 646, 49–60. doi:[10.1086/504825](https://doi.org/10.1086/504825).
- Rodriguez-Gomez, V.; Genel, S.; Vogelsberger, M.; Sijacki, D.; Pillepich, A.; Sales, L.V.; Torrey, P.; Snyder, G.; Nelson, D.; Springel, V.; et al. (2015). The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models. *Mon. Not. R. Astron. Soc.*, 449, 49–64. doi:[10.1093/mnras/stv264](https://doi.org/10.1093/mnras/stv264).
- Roebber, E.; Holder, G.; Holz, D.E.; Warren, M. (2016). Cosmic Variance in the Nanohertz Gravitational

- Wave Background. *Astrophys. J.*, 819, 163. doi:[10.3847/0004-637X/819/2/163](https://doi.org/10.3847/0004-637X/819/2/163).
- Roedig, C.; Dotti, M.; Sesana, A.; Cuadra, J.; Colpi, M. (2011). Limiting eccentricity of subparsec massive black hole binaries surrounded by self-gravitating gas discs. *Mon. Not. R. Astron. Soc.*, 415, 3033–3041. doi:[10.1111/j.1365-2966.2011.18927.x](https://doi.org/10.1111/j.1365-2966.2011.18927.x).
- Roedig, C.; Sesana, A.; Dotti, M.; Cuadra, J.; Amaro-Seoane, P.; Haardt, F. (2012). Evolution of binary black holes in self gravitating discs. Dissecting the torques. *Astron. Astrophys.*, 545, A127. doi:[10.1051/0004-6361/201219986](https://doi.org/10.1051/0004-6361/201219986).
- Romani, R.W.; Taylor, J.H. (1983). An upper limit on the stochastic background of ultralow-frequency gravitational waves. *Astrophys. J. Lett.*, 265, L35–L37. doi:[10.1086/183953](https://doi.org/10.1086/183953).
- Roos, N.; Kaastra, J.S.; Hummel, C.A. (1993). A massive binary black hole in 1928 + 738? *Astrophys. J.*, 409, 130–133. doi:[10.1086/172647](https://doi.org/10.1086/172647).
- Rosado, P.A.; Sesana, A.; Gair, J. (2015). Expected properties of the first gravitational wave signal detected with pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 451, 2417–2433. doi:[10.1093/mnras/stv1098](https://doi.org/10.1093/mnras/stv1098).
- Ryu, D.; Ostriker, J.P.; Kang, H.; Cen, R. (1993). A cosmological hydrodynamic code based on the total variation diminishing scheme. *Astrophys. J.*, 414, 1–19. doi:[10.1086/173051](https://doi.org/10.1086/173051).
- Ryu, T.; Perna, R.; Haiman, Z.; Ostriker, J.P.; Stone, N.C. (2018). Interactions between multiple supermassive black holes in galactic nuclei: a solution to the final parsec problem. *Mon. Not. R. Astron. Soc.*, 473, 3410–3433. doi:[10.1093/mnras/stx2524](https://doi.org/10.1093/mnras/stx2524).
- Salpeter, E.E. (1964). Accretion of Interstellar Matter by Massive Objects. *Astrophys. J.*, 140, 796–800. doi:[10.1086/147973](https://doi.org/10.1086/147973).
- Sampson, L.; Cornish, N.J.; McWilliams, S.T. (2015). Constraining the solution to the last parsec problem with pulsar timing. *Phys. Rev. D*, 91, 8, 084055. doi:[10.1103/PhysRevD.91.084055](https://doi.org/10.1103/PhysRevD.91.084055).
- Sanders, D.B.; Soifer, B.T.; Elias, J.H.; Madore, B.F.; Matthews, K.; Neugebauer, G.; Scoville, N.Z. (1988). Ultraluminous infrared galaxies and the origin of quasars. *Astrophys. J.*, 325, 74–91. doi:[10.1086/165983](https://doi.org/10.1086/165983).
- Sarajedini, V.L.; Gilliland, R.L.; Kasm, C. (2003). A V-Band Survey for Variable Galactic Nuclei in the Hubble Deep Field. *Astrophys. J.*, 599, 173–184. doi:[10.1086/379168](https://doi.org/10.1086/379168).
- Sazhin, M.V. (1978). Opportunities for detecting ultralong gravitational waves. *Soviet Astronomy*, 22, 36–38.
- Sądowski, A.; Narayan, R.; Tchekhovskoy, A.; Abarca, D.; Zhu, Y.; McKinney, J.C. (2015). Global simulations of axisymmetric radiative black hole accretion discs in general relativity with a mean-field magnetic dynamo. *Mon. Not. R. Astron. Soc.*, 447, 49–71. doi:[10.1093/mnras/stu2387](https://doi.org/10.1093/mnras/stu2387).
- Schutz, K.; Ma, C.P. (2016). Constraints on individual supermassive black hole binaries from pulsar timing array limits on continuous gravitational waves. *Mon. Not. R. Astron. Soc.*, 459, 1737–1744. doi:[10.1093/mnras/stw768](https://doi.org/10.1093/mnras/stw768).



- Sellwood, J.A.; Wilkinson, A. (1993). Dynamics of barred galaxies. *Reports on Progress in Physics*, 56, 173–256. doi:[10.1088/0034-4885/56/2/001](https://doi.org/10.1088/0034-4885/56/2/001).
- Sesana, A. (2010). Self Consistent Model for the Evolution of Eccentric Massive Black Hole Binaries in Stellar Environments: Implications for Gravitational Wave Observations. *Astrophys. J.*, 719, 851–864. doi:[10.1088/0004-637X/719/1/851](https://doi.org/10.1088/0004-637X/719/1/851).
- Sesana, A. (2013a). Insights into the astrophysics of supermassive black hole binaries from pulsar timing observations. *Classical and Quantum Gravity*, 30, 22, 224014. doi:[10.1088/0264-9381/30/22/224014](https://doi.org/10.1088/0264-9381/30/22/224014).
- Sesana, A. (2013b). Systematic investigation of the expected gravitational wave signal from supermassive black hole binaries in the pulsar timing band. *Mon. Not. R. Astron. Soc.*, 433, L1–L5. doi:[10.1093/mnrasl/slt034](https://doi.org/10.1093/mnrasl/slt034).
- Sesana, A.; Gualandris, A.; Dotti, M. (2011). Massive black hole binary eccentricity in rotating stellar systems. *Mon. Not. R. Astron. Soc.*, 415, L35–L39. doi:[10.1111/j.1745-3933.2011.01073.x](https://doi.org/10.1111/j.1745-3933.2011.01073.x).
- Sesana, A.; Haardt, F.; Madau, P. (2006). Interaction of Massive Black Hole Binaries with Their Stellar Environment. I. Ejection of Hypervelocity Stars. *Astrophys. J.*, 651, 392–400. doi:[10.1086/507596](https://doi.org/10.1086/507596).
- Sesana, A.; Haardt, F.; Madau, P. (2007). Interaction of Massive Black Hole Binaries with Their Stellar Environment. II. Loss Cone Depletion and Binary Orbital Decay. *Astrophys. J.*, 660, 546–555. doi:[10.1086/513016](https://doi.org/10.1086/513016).
- Sesana, A.; Haardt, F.; Madau, P.; Volonteri, M. (2004). Low-Frequency Gravitational Radiation from Coalescing Massive Black Hole Binaries in Hierarchical Cosmologies. *Astrophys. J.*, 611, 623–632. doi:[10.1086/422185](https://doi.org/10.1086/422185).
- Sesana, A.; Haiman, Z.; Kocsis, B.; Kelley, L.Z. (2017a). Testing the binary hypothesis: pulsar timing constraints on supermassive black hole binary candidates. *ArXiv e-prints*.
- Sesana, A.; Haiman, Z.; Kocsis, B.; Kelley, L.Z. (2017b). Testing the binary hypothesis: pulsar timing constraints on supermassive black hole binary candidates. *ArXiv e-prints*.
- Sesana, A.; Khan, F.M. (2015). Scattering experiments meet N-body - I. A practical recipe for the evolution of massive black hole binaries in stellar environments. *Mon. Not. R. Astron. Soc.*, 454, L66–L70. doi:[10.1093/mnrasl/slv131](https://doi.org/10.1093/mnrasl/slv131).
- Sesana, A.; Roedig, C.; Reynolds, M.T.; Dotti, M. (2012). Multimessenger astronomy with pulsar timing and X-ray observations of massive black hole binaries. *Mon. Not. R. Astron. Soc.*, 420, 860–877. doi:[10.1111/j.1365-2966.2011.20097.x](https://doi.org/10.1111/j.1365-2966.2011.20097.x).
- Sesana, A.; Shankar, F.; Bernardi, M.; Sheth, R.K. (2016). Selection bias in dynamically measured supermassive black hole samples: consequences for pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 463, L6–L11. doi:[10.1093/mnrasl/slw139](https://doi.org/10.1093/mnrasl/slw139).
- Sesana, A.; Vecchio, A.; Colacino, C.N. (2008). The stochastic gravitational-wave background from massive black hole binary systems: implications for observations with Pulsar Timing Arrays. *Mon. Not. R. Astron. Soc.*, 390, 192–209. doi:[10.1111/j.1365-2966.2008.13682.x](https://doi.org/10.1111/j.1365-2966.2008.13682.x).
- Sesana, A.; Vecchio, A.; Volonteri, M. (2009). Gravitational waves from resolvable massive black hole bi-

- nary systems and observations with Pulsar Timing Arrays. *Mon. Not. R. Astron. Soc.*, 394, 2255–2265. doi:[10.1111/j.1365-2966.2009.14499.x](https://doi.org/10.1111/j.1365-2966.2009.14499.x).
- Shakura, N.I.; Sunyaev, R.A. (1973). Black holes in binary systems. Observational appearance. *Astron. Astrophys.*, 24, 337–355.
- Shankar, F.; Bernardi, M.; Sheth, R.K.; Ferrarese, L.; Graham, A.W.; Savorgnan, G.; Allevato, V.; Marconi, A.; Läsker, R.; Lapi, A. (2016). Selection bias in dynamically measured supermassive black hole samples: its consequences and the quest for the most fundamental relation. *Mon. Not. R. Astron. Soc.*, 460, 3119–3142. doi:[10.1093/mnras/stw678](https://doi.org/10.1093/mnras/stw678).
- Shannon, R.M.; Ravi, V.; Lentati, L.T.; Lasky, P.D.; Hobbs, G.; Kerr, M.; Manchester, R.N.; Coles, W.A.; Levin, Y.; Bailes, M.; et al. (2015). Gravitational waves from binary supermassive black holes missing in pulsar observations. *Science*, 349, 1522–1525. doi:[10.1126/science.aab1910](https://doi.org/10.1126/science.aab1910).
- Shapiro, S.L.; Teukolsky, S.A. (1986). *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects*.
- Shen, Y.; Greene, J.E.; Strauss, M.A.; Richards, G.T.; Schneider, D.P. (2008). Biases in Virial Black Hole Masses: An SDSS Perspective. *Astrophys. J.*, 680, 169–190. doi:[10.1086/587475](https://doi.org/10.1086/587475).
- Shen, Y.; Liu, X.; Greene, J.E.; Strauss, M.A. (2011). Type 2 Active Galactic Nuclei with Double-peaked [O III] Lines. II. Single AGNs with Complex Narrow-line Region Kinematics are More Common than Binary AGNs. *Astrophys. J.*, 735, 48. doi:[10.1088/0004-637X/735/1/48](https://doi.org/10.1088/0004-637X/735/1/48).
- Sijacki, D.; Springel, V.; Di Matteo, T.; Hernquist, L. (2007). A unified model for AGN feedback in cosmological simulations of structure formation. *Mon. Not. R. Astron. Soc.*, 380, 877–900. doi:[10.1111/j.1365-2966.2007.12153.x](https://doi.org/10.1111/j.1365-2966.2007.12153.x).
- Sijacki, D.; Vogelsberger, M.; Genel, S.; Springel, V.; Torrey, P.; Snyder, G.F.; Nelson, D.; Hernquist, L. (2015). The Illustris simulation: the evolving population of black holes across cosmic time. *Mon. Not. R. Astron. Soc.*, 452, 575–596. doi:[10.1093/mnras/stv1340](https://doi.org/10.1093/mnras/stv1340).
- Simon, J.; Burke-Spolaor, S. (2016). Constraints on Black Hole/Host Galaxy Co-evolution and Binary Stalling Using Pulsar Timing Arrays. *Astrophys. J.*, 826, 11. doi:[10.3847/0004-637X/826/1/11](https://doi.org/10.3847/0004-637X/826/1/11).
- Small, T.A.; Blandford, R.D. (1992). Quasar evolution and the growth of black holes. *Mon. Not. R. Astron. Soc.*, 259, 725–737. doi:[10.1093/mnras/259.4.725](https://doi.org/10.1093/mnras/259.4.725).
- Snyder, G.F.; Torrey, P.; Lotz, J.M.; Genel, S.; McBride, C.K.; Vogelsberger, M.; Pillepich, A.; Nelson, D.; Sales, L.V.; Sijacki, D.; et al. (2015). Galaxy morphology and star formation in the Illustris Simulation at  $z = 0$ . *Mon. Not. R. Astron. Soc.*, 454, 1886–1908. doi:[10.1093/mnras/stv2078](https://doi.org/10.1093/mnras/stv2078).
- Soltan, A. (1982). Masses of quasars. *Mon. Not. R. Astron. Soc.*, 200, 115–122. doi:[10.1093/mnras/200.1.115](https://doi.org/10.1093/mnras/200.1.115).
- Springel, V. (2005). The cosmological simulation code GADGET-2. *Mon. Not. R. Astron. Soc.*, 364, 1105–1134. doi:[10.1111/j.1365-2966.2005.09655.x](https://doi.org/10.1111/j.1365-2966.2005.09655.x).
- Springel, V. (2010). E pur si muove: Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Mon. Not. R. Astron. Soc.*, 401, 791–851. doi:[10.1111/j.1365-2966.2009.15715.x](https://doi.org/10.1111/j.1365-2966.2009.15715.x).

- Springel, V.; Hernquist, L. (2002). Cosmological smoothed particle hydrodynamics simulations: the entropy equation. *Mon. Not. R. Astron. Soc.*, 333, 649–664. doi:[10.1046/j.1365-8711.2002.05445.x](https://doi.org/10.1046/j.1365-8711.2002.05445.x).
- Springel, V.; Hernquist, L. (2003). Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *Mon. Not. R. Astron. Soc.*, 339, 289–311. doi:[10.1046/j.1365-8711.2003.06206.x](https://doi.org/10.1046/j.1365-8711.2003.06206.x).
- Springel, V.; White, S.D.M.; Jenkins, A.; Frenk, C.S.; Yoshida, N.; Gao, L.; Navarro, J.; Thacker, R.; Croton, D.; Helly, J.; et al. (2005). Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435, 629–636. doi:[10.1038/nature03597](https://doi.org/10.1038/nature03597).
- Tanaka, T.; Menou, K.; Haiman, Z. (2012). Electromagnetic counterparts of supermassive black hole binaries resolved by pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 420, 705–719. doi:[10.1111/j.1365-2966.2011.20083.x](https://doi.org/10.1111/j.1365-2966.2011.20083.x).
- Tang, Y.; MacFadyen, A.; Haiman, Z. (2017). On the orbital evolution of supermassive black hole binaries with circumbinary accretion discs. *Mon. Not. R. Astron. Soc.*, 469, 4258–4267. doi:[10.1093/mnras/stx1130](https://doi.org/10.1093/mnras/stx1130).
- Taylor, S.; Ellis, J.; Gair, J. (2014). Accelerated Bayesian model-selection and parameter-estimation in continuous gravitational-wave searches with pulsar-timing arrays. *Phys. Rev. D*, 90, 10, 104028. doi:[10.1103/PhysRevD.90.104028](https://doi.org/10.1103/PhysRevD.90.104028).
- Taylor, S.R.; Gair, J.R. (2013). Searching for anisotropic gravitational-wave backgrounds using pulsar timing arrays. *Phys. Rev. D*, 88, 8, 084001. doi:[10.1103/PhysRevD.88.084001](https://doi.org/10.1103/PhysRevD.88.084001).
- Taylor, S.R.; Huerta, E.A.; Gair, J.R.; McWilliams, S.T. (2016a). Detecting Eccentric Supermassive Black Hole Binaries with Pulsar Timing Arrays: Resolvable Source Strategies. *Astrophys. J.*, 817, 70. doi:[10.3847/0004-637X/817/1/70](https://doi.org/10.3847/0004-637X/817/1/70).
- Taylor, S.R.; Mingarelli, C.M.F.; Gair, J.R.; Sesana, A.; Theureau, G.; Babak, S.; Bassa, C.G.; Brem, P.; Burgay, M.; Caballero, R.N.; et al. (2015). Limits on Anisotropy in the Nanohertz Stochastic Gravitational Wave Background. *Physical Review Letters*, 115, 4, 041101. doi:[10.1103/PhysRevLett.115.041101](https://doi.org/10.1103/PhysRevLett.115.041101).
- Taylor, S.R.; Simon, J.; Sampson, L. (2017). Constraints on the Dynamical Environments of Supermassive Black-Hole Binaries Using Pulsar-Timing Arrays. *Physical Review Letters*, 118, 18, 181102. doi:[10.1103/PhysRevLett.118.181102](https://doi.org/10.1103/PhysRevLett.118.181102).
- Taylor, S.R.; Vallisneri, M.; Ellis, J.A.; Mingarelli, C.M.F.; Lazio, T.J.W.; van Haasteren, R. (2016b). Are We There Yet? Time to Detection of Nanohertz Gravitational Waves Based on Pulsar-timing Array Limits. *Astrophys. J. Lett.*, 819, L6. doi:[10.3847/2041-8205/819/1/L6](https://doi.org/10.3847/2041-8205/819/1/L6).
- The eLISA Consortium; Seoane, P.A.; Aoudia, S.; Audley, H.; Auger, G.; Babak, S.; Baker, J.; Barausse, E.; Barke, S.; et al. (2013). The Gravitational Universe. *ArXiv e-prints*.
- The LIGO Scientific Collaboration; the Virgo Collaboration (2016). Observation of Gravitational Waves from a Binary Black Hole Merger. *ArXiv e-prints*.
- The NANOGrav Collaboration; Arzoumanian, Z.; Brazier, A.; Burke-Spolaor, S.; Chamberlin, S.; Chatterjee, S.; Christy, B.; Cordes, J.M.; Cornish, N.; Crowter, K.; et al. (2015). The NANOGrav Nine-year

- Data Set: Observations, Arrival Time Measurements, and Analysis of 37 Millisecond Pulsars. *Astrophys. J.*, 813, 65. doi:[10.1088/0004-637X/813/1/65](https://doi.org/10.1088/0004-637X/813/1/65).
- Torrey, P.; Vogelsberger, M.; Genel, S.; Sijacki, D.; Springel, V.; Hernquist, L. (2014). A model for cosmological simulations of galaxy formation physics: multi-epoch validation. *Mon. Not. R. Astron. Soc.*, 438, 1985–2004. doi:[10.1093/mnras/stt2295](https://doi.org/10.1093/mnras/stt2295).
- Tsalmantza, P.; Decarli, R.; Dotti, M.; Hogg, D.W. (2011). A Systematic Search for Massive Black Hole Binaries in the Sloan Digital Sky Survey Spectroscopic Sample. *Astrophys. J.*, 738, 20. doi:[10.1088/0004-637X/738/1/20](https://doi.org/10.1088/0004-637X/738/1/20).
- Urry, C.M.; Padovani, P. (1995). Unified Schemes for Radio-Loud Active Galactic Nuclei. *PASP*, 107, 803. doi:[10.1086/133630](https://doi.org/10.1086/133630).
- Valtonen, M.J.; Lehto, H.J.; Nilsson, K.; Heidt, J.; Takalo, L.O.; Sillanpää, A.; Villforth, C.; Kidger, M.; Poyner, G.; Pursimo, T.; et al. (2008). A massive binary black-hole system in OJ287 and a test of general relativity. *Nature*, 452, 851–853. doi:[10.1038/nature06896](https://doi.org/10.1038/nature06896).
- van den Bosch, F.C.; Lewis, G.F.; Lake, G.; Stadel, J. (1999). Substructure in Dark Halos: Orbital Eccentricities and Dynamical Friction. *Astrophys. J.*, 515, 50–68. doi:[10.1086/307023](https://doi.org/10.1086/307023).
- van der Walt, S.; Colbert, S.C.; Varoquaux, G. (2011). The numpy array: a structure for efficient numerical computation. *CoRR*, 1102.1523.
- Vasiliev, E.; Antonini, F.; Merritt, D. (2014). The Final-parsec Problem in Nonspherical Galaxies Revisited. *Astrophys. J.*, 785, 163. doi:[10.1088/0004-637X/785/2/163](https://doi.org/10.1088/0004-637X/785/2/163).
- Vasiliev, E.; Antonini, F.; Merritt, D. (2015). The Final-parsec Problem in the Collisionless Limit. *Astrophys. J.*, 810, 49. doi:[10.1088/0004-637X/810/1/49](https://doi.org/10.1088/0004-637X/810/1/49).
- Verbiest, J.P.W.; Lentati, L.; Hobbs, G.; van Haasteren, R.; Demorest, P.B.; Janssen, G.H.; Wang, J.B.; Desvignes, G.; Caballero, R.N.; Keith, M.J.; et al. (2016). The International Pulsar Timing Array: First data release. *Mon. Not. R. Astron. Soc.*, 458, 1267–1288. doi:[10.1093/mnras/stw347](https://doi.org/10.1093/mnras/stw347).
- Vigeland, S.J.; Siemens, X. (2016). Supermassive black hole binary environments: Effects on the scaling laws and time to detection for the stochastic background. *Phys. Rev. D*, 94, 12, 123003. doi:[10.1103/PhysRevD.94.123003](https://doi.org/10.1103/PhysRevD.94.123003).
- Vogelsberger, M.; Genel, S.; Sijacki, D.; Torrey, P.; Springel, V.; Hernquist, L. (2013). A model for cosmological simulations of galaxy formation physics. *Mon. Not. R. Astron. Soc.*, 436, 3031–3067. doi:[10.1093/mnras/stt1789](https://doi.org/10.1093/mnras/stt1789).
- Vogelsberger, M.; Genel, S.; Springel, V.; Torrey, P.; Sijacki, D.; Xu, D.; Snyder, G.; Bird, S.; Nelson, D.; Hernquist, L. (2014a). Properties of galaxies reproduced by a hydrodynamic simulation. *Nature*, 509, 177–182. doi:[10.1038/nature13316](https://doi.org/10.1038/nature13316).
- Vogelsberger, M.; Genel, S.; Springel, V.; Torrey, P.; Sijacki, D.; Xu, D.; Snyder, G.; Nelson, D.; Hernquist, L. (2014b). Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *Mon. Not. R. Astron. Soc.*, 444, 1518–1547. doi:[10.1093/mnras/stu1536](https://doi.org/10.1093/mnras/stu1536).
- Volonteri, M.; Madau, P.; Haardt, F. (2003). The Formation of Galaxy Stellar Cores by the Hierarchical

- Merging of Supermassive Black Holes. *Astrophys. J.*, 593, 661–666. doi:[10.1086/376722](https://doi.org/10.1086/376722).
- Wang, J.M.; Chen, Y.M.; Hu, C.; Mao, W.M.; Zhang, S.; Bian, W.H. (2009). Active Galactic Nuclei with Double-Peaked Narrow Lines: Are they Dual Active Galactic Nuclei? *Astrophys. J. Lett.*, 705, L76–L80. doi:[10.1088/0004-637X/705/1/L76](https://doi.org/10.1088/0004-637X/705/1/L76).
- White, S.D.M.; Frenk, C.S. (1991). Galaxy formation through hierarchical clustering. *Astrophys. J.*, 379, 52–79. doi:[10.1086/170483](https://doi.org/10.1086/170483).
- Wong, K.W.; Irwin, J.A.; Yukita, M.; Million, E.T.; Mathews, W.G.; Bregman, J.N. (2011). Resolving the Bondi Accretion Flow toward the Supermassive Black Hole of NGC 3115 with Chandra. *Astrophys. J. Lett.*, 736, L23. doi:[10.1088/2041-8205/736/1/L23](https://doi.org/10.1088/2041-8205/736/1/L23).
- Wyithe, J.S.B.; Loeb, A. (2003). Low-Frequency Gravitational Waves from Massive Black Hole Binaries: Predictions for LISA and Pulsar Timing Arrays. *Astrophys. J.*, 590, 691–706. doi:[10.1086/375187](https://doi.org/10.1086/375187).
- Xie, G.Z.; Liang, E.W.; Zhou, S.B.; Li, K.H.; Dai, B.Z.; Ma, L. (2002). Photometry of three gamma-ray-loud quasars and implications for supermassive black holes\*. *Mon. Not. R. Astron. Soc.*, 334, 459–470. doi:[10.1046/j.1365-8711.2002.05528.x](https://doi.org/10.1046/j.1365-8711.2002.05528.x).
- Yardley, D.R.B.; Coles, W.A.; Hobbs, G.B.; Verbiest, J.P.W.; Manchester, R.N.; van Straten, W.; Jenet, F.A.; Bailes, M.; Bhat, N.D.R.; Burke-Spolaor, S.; et al. (2011). On detection of the stochastic gravitational-wave background using the Parkes pulsar timing array. *Mon. Not. R. Astron. Soc.*, 414, 1777–1787. doi:[10.1111/j.1365-2966.2011.18517.x](https://doi.org/10.1111/j.1365-2966.2011.18517.x).
- Young, J.S.; Scoville, N.Z. (1991). Molecular gas in galaxies. *Annu. Rev. Astron. Astrophys.*, 29, 581–625. doi:[10.1146/annurev.aa.29.090191.003053](https://doi.org/10.1146/annurev.aa.29.090191.003053).
- Yu, Q. (2002). Evolution of massive binary black holes. *Mon. Not. R. Astron. Soc.*, 331, 935–958. doi:[10.1046/j.1365-8711.2002.05242.x](https://doi.org/10.1046/j.1365-8711.2002.05242.x).
- Yu, Q.; Tremaine, S. (2002). Observational constraints on growth of massive black holes. *Mon. Not. R. Astron. Soc.*, 335, 965–976. doi:[10.1046/j.1365-8711.2002.05532.x](https://doi.org/10.1046/j.1365-8711.2002.05532.x).
- Zhu, X.J.; Hobbs, G.; Wen, L.; Coles, W.A.; Wang, J.B.; Shannon, R.M.; Manchester, R.N.; Bailes, M.; Bhat, N.D.R.; Burke-Spolaor, S.; et al. (2014). An all-sky search for continuous gravitational waves in the Parkes Pulsar Timing Array data set. *Mon. Not. R. Astron. Soc.*, 444, 3709–3720. doi:[10.1093/mnras/stu1717](https://doi.org/10.1093/mnras/stu1717).
- Zhu, X.J.; Wen, L.; Hobbs, G.; Zhang, Y.; Wang, Y.; Madison, D.R.; Manchester, R.N.; Kerr, M.; Rosado, P.A.; Wang, J.B. (2015). Detection and localization of single-source gravitational waves with pulsar timing arrays. *Mon. Not. R. Astron. Soc.*, 449, 1650–1663. doi:[10.1093/mnras/stv381](https://doi.org/10.1093/mnras/stv381).
- Zhu, X.J.; Wen, L.; Xiong, J.; Xu, Y.; Wang, Y.; Mohanty, S.D.; Hobbs, G.; Manchester, R.N. (2016). Detection and localization of continuous gravitational waves with pulsar timing arrays: the role of pulsar terms. *Mon. Not. R. Astron. Soc.*, 461, 1317–1327. doi:[10.1093/mnras/stw1446](https://doi.org/10.1093/mnras/stw1446).
- Zwicky, F. (1956). Multiple Galaxies. *Ergebnisse der exakten Naturwissenschaften*, 29, 344–385.