Study of formation of binary black holes, their interactions with surroundings and their mergers

A Dissertation presented

by

Taeho Ryu

to

The Graduate School

in Partial Fulfillment of the

Requirements

for the Degree of

Doctor of Philosophy

in

Physics

Stony Brook University

May 2018

Stony Brook University

The Graduate School

Taeho Ryu

We, the dissertation committe for the above candidate for the

Doctor of Philosophy degree, hereby recommend

acceptance of this dissertation

Rosalba Perna - Dissertation Advisor
Professor, Department of Physics & Astronomy

Jin Koda - Chairperson of Defense
Associate Professor, Department of Physics & Astronomy

Matthew Dawber - Committee Member
Associate Professor, Department of Physics & Astronomy

Yuri Levin - External Committee Member
Professor, Columbia University

This dissertation is accepted by the Graduate School

Charles Taber
Dean of the Graduate School

Abstract of the Dissertation

Study of formation of binary black holes, their interactions with surroundings and their mergers

by

Taeho Ryu

Doctor of Philosophy

in

Physics

Stony Brook University

2018

Black holes have existed since the collapse of the first stars in the early Universe, and play a fundamental role in its evolution. In particular, binary black holes are a natural outcome as a result of dynamical interactions among themselves and with surrounding stars. Goal of this dissertation is to investigate the formation of binary black holes, their dynamics and mergers, with a focus on their global-scale observational imprints, from low to high redshift. To this aim, we developed a multi-purpose N-body code. The code is designed to include many physical effects relevant for each different model, including gravitational attraction up to 2.5-order post-Newtonian terms, dynamical friction, force from a static or dynamically evolving background potential, mass growth via accretion. This code has been used to perform the studies presented in this dissertation. We provide a full description of the code in §2. We present a study of the formation of Population III X-ray binaries and their X-ray output in §3. Using N-body simulations, we estimate their formation rates and hence discuss the effects of their X-ray heating on the early intergalactic medium, along with the observational implications for the 21 cm line detection. In §4, we investigate the growth of Population III remnant black holes to intermediate-mass and supermassive black holes via accretion at a very rapid rate ("hyper-Eddington accretion"). We suggest this rapid accretion as a viable mechanism for the formation of massive black holes at the galactic nuclei. Furthermore, we find that natural outcomes of the rapidly growing black holes are extreme mass ratio inspirals (EMRI), one of the main gravitational wave sources of the planned *eLISA*. Finally, we study the evolution of supermassive black hole binaries in massive galaxies at $0 < z < 4$ in §5. We model their mergers in a dynamically evolving galaxy model and infer their merger rates and the stochastic gravitational wave background from their mergers. We also compute the shape of the gravitational wave signal and the effects of high eccentricities on the spectrum, which is relevant for future detections with the Pulsar Timing Array (PTA).

Dedication Page

To my family, my loving nieces, Jiyul and Chaeyoon and my nephew, Rone.

To the memory of professor Moo Young Han.
He gave me the reason to come to the US for higher studies in physics.

Contents

List of Figures

List of Tables

Acknowledgements

I am grateful to my advisor, Rosalba Perna, for her invaluable advice and physical insights. Contacting her when I got into this PhD program was one of my best decisions that I made during my entire PhD year. I still clearly remember the moment when I had the very first meeting with her in her office. Above all, I can not thank more for her unceasing encouragement. Whatever I tried to do, she always listened to me. Before I started working with her, there were some times when I had lost confidence in myself after unsuccessful journeys to search for fields of physics that interested me and I may be good at. Her encouragement and supports played a pivotal role in getting me motivated again. So I could embark on this new journey. Her wise advice and suggestions have always been good guidances for me to move forward whenever I made mistakes and encountered challenges to overcome and help me finish this journey, and more importantly, will be in my future path.

I thank my committee members, Jin Koda, Matthew Dawber and Yuri Levin, for their constructive comments which helped me to improve this dissertation. I learnt even during my defense presentation from them what I should keep in mind as an astrophysicist and what physics I should pursue.

I wish to thank all of my collaborators, Takamitsu Tanaka, Nathan Leigh, Nicholas Stone, Krzysztof Belczynski, Zoltan Haiman and Jeremiah Ostriker for their meaningful comments and constructive feedback. This work could not have been done without their helps.

And I thank my best friends, Woosang Chung, for supporting me in many ways. Even a short chat with him gave me reasons and motivations to live harder. I thank Moshe Kellerstein for helping me prepare my defense presentation as well as other presentations. Without his helps, my defense presentation would have been much less successful. I sincerely hope his future journey will be bright and be filled with happiness. I thank my friends, Donald Wilcox, Maria Barrios Sazo, Tianqi Zhao, Mu-Hung Chang, Dongwon Han, Yihan Wang, in SBU astronomy group. It was always enjoyful to have a chat with them. Finally, I also thank my friends, Gongjun Choi, Joonhyuk Kwon, Jun-Sik Yoo, Youngshin Kim, Saebyeok Jeong, HyeYun Park, Jae Hyeok Chang, Nanoom Lee, Kyle Lee, Changhyun Cho, Chang Ha Choi. They came to my defense presentation and showed their supports to me. I really appreciate it.

I want to express my sincere gratitude to all for finishing this journey.

Publications

Articles in refereed Journals

8. Timur Ibragimov, Nathan W. C. Leigh, Taeho Ryu, Teresa Panurach, Rosalba Perna

   When do star clusters become multiple star systems? II. Toward a half-life formalism with four bodies, accepted for publication in MNRAS (2018)

7. Taeho Ryu, Rosalba Perna, Zóltan Haiman, Jeremiah P. Ostriker, Nicholas. C. Stone

   Interactions between multiple supermassive black holes in galactic nuclei: a solution to the final parsec problem, MNRAS, 473, 3410 (2018)

6. Krzysztof Belczynski, Taeho Ryu, Rosalba Perna, Emanuele Berti, Takamitsu L. Tanaka, Tomasz Bulik

   On the likelihood of detecting gravitational waves from Population III compact object binaries, MNRAS, 471, 4702 (2017)

5. Taeho Ryu, Nathan W. C. Leigh, Rosalba Perna

   Formation of runaway stars in a star-cluster potential, MNRAS, 470, 3049 (2017)

4. Taeho Ryu, Nathan W. C. Leigh, Rosalba Perna

   An analytic method for identifying dynamically-formed runaway stars, MNRAS, 470, 2 (2017)

3. Taeho Ryu, Nathan W. C. Leigh, Rosalba Perna

   Numerical study of the $N = 4$ binary-binary scatterings in a background potential, MNRAS, 467, 4447 (2017)

2. Taeho Ryu, Takamitsu L. Tanaka, Rosalba Perna, Zóltan Haiman

   Intermediate-mass black holes from Population III remnants in the first galactic nuclei, MNRAS, 460, 4122 (2016)

1. Taeho Ryu, Takamitsu L. Tanaka, Rosalba Perna

   Formation, disruption and energy output of Population III X-ray binaries, MNRAS, 456, 223 (2016)

Articles in conference proceedings

1. Taeho Ryu, Tanaka, T. L., Perna, R. *Population III X-Ray Binaries*, in "Frontier Research in Astrophysics − II", Italy, (2016).

1 Introduction

Black holes have played a pivotal role in shaping the Universe, from small to large scales since the collapse of the first stars in the early Universe. In this dissertation, we study the formation of binary BHs, their dynamics and their observational imprints (X-ray and gravitational wave emission), from stellar-mass BH binaries to supermassive BH binaries across the history of the Universe.

1.1 The first star remnant black holes and their binary formation

1.1.1 Multiplicity of the first (Pop III) stars and the primordial initial mass function

In the standard ΛCDM cosmological model, large structures are built hierarchically via mergers of smaller objects. Those smaller structures, which consist of dark matter (DM), form as a result of gravitational instabilities from small fluctuations of the primordial density field. These fluctuations are further amplified by gravity. Once the mass of the gravitationally bound DM object becomes larger than the cosmological Jeans mass $M_{\rm J}$, pressure gradients can no longer balance against the gravity of the DM. This results in collapse of the DM along with infall of gas. When a virial equilibrium state with a virial temperature $T_{\rm vir} < 10^4$ K is reached, a "mini dark matter halo" is formed. The first stars, also known as population III stars or pop III stars, are expected to form in these first mini DM halos with $M_{\rm DM} \simeq 10^{5-6}$ $M_\odot$ at $z \sim 20 - 50$ where cooling through ro-vibrational transitions of hydrogen molecules H_2 is effective (Haiman et al., 1996; Tegmark et al., 1997). The characteristic mass of the DM halo with sufficient amount of H_2 is determined by the condition that the cooling time scale is shorter than the dynamical time scale (Rees & Ostriker, 1977; Silk, 1977). See Figure 1 of Bromm (2013).

The density of the gas keeps increasing until most of the hydrogen molecules are destroyed by collisions ($n > 10^{16}$ cm^{-3}, where n is the number density). The gas becomes adiabatic and this point is typically identified as the formation of protostars. It has been found in numerical simulations that the mass of the protostars at their formation is around $10^{-2} - 10^{-1}$ $M_\odot$, but subsequently they grow rapidly via accretion at a rate of $\dot{M}_{\rm acc} \sim 10^{-2}$ $M_\odot$ yr^{-1} (Abel et al., 2002). [1]

If no fragmentation occurs so that only one protostar forms and subsequently grows at such a high rate, it is expected that the mass of the first stars would reach more than 100 $M_\odot$. This picture of "a single massive star per halo" was suggested in numerical simulations (e.g.

[1] Such high accretion rate can be understood by estimating the rate at which a Jeans mass accretes on the protostar on its free-fall time scale $t_{\rm ff}$ (Shu, 1977),

$$\dot{M}_{\rm acc} \simeq \frac{M_{\rm J}}{t_{\rm ff}} \simeq \frac{c_{\rm s}^3}{G} \propto T^{3/2}. \quad (1)$$

Considering higher temperature of the gas cloud in the mini halo ($T > 200$ K, a characteristic temperature with $n \sim 10^4$ cm^{-3} for efficient H_2 cooling) than that of star forming gas clouds in the local universe, typical accretion rates in the mini halo are higher roughly by two orders of magnitude that those in nearby star forming gas clouds.

Abel et al., 2000, 2002; Bromm et al., 2002). If such massive stars in the halos survive until death, one can expect that the masses of the remnant BHs are also very high. Generally speaking, black holes can form from the collapse of non-rotating, metal-free star with masses between $\sim 25 - 140$ $M_\odot$ and > 260 $M_\odot$ (Heger et al., 2003; Zhang et al., 2008). $25 - 40$ $M_\odot$ stars go though significant mass losses when they collapse, so the BH remnants end up with $\sim 40\%$ of the progenitor mass (Zhang et al., 2008). The stars within the rest of the mass range (i.e. 40 $M_\odot - 140$ $M_\odot$ or > 260 $M_\odot$) directly collapse to a BH with a mass comparable to that of its progenitor star. On the other hand, stars with masses outside of this range (i.e., ≤ 25 $M_\odot$ and $140 - 260$ $M_\odot$) do not leave behind a BH remnant.

Within this "single massive star" picture, it appears that the likelihood of formation of binaries is inevitably low unless fragmentation occurs. These studies find that instabilities can not lead to fragmentation of the gas. The chemothermal instabilities are suppressed since the crossing times in the collapsing core are short. In other words, media with different temperatures are efficiently mixed so the fluctuations never grow significantly (Abel et al., 2002; Yoshida et al., 2006; Turk et al., 2009). Furthermore, Abel et al. (2002) claim that rotational support is not strong enough to halt the collapse, based on initially small angular momentum of the infalling gas compared to the mean value ($\simeq 0.1\%$) and effective angular momentum transport by shock waves.

However, recent studies, using improved hydrodynamics simulations with higher resolution, suggest that fragmentation of gas can continue to occur even in the presence of the star which first formed prior to the others (Turk et al., 2009; Stacy et al., 2010; Greif et al., 2012). Contrary to the findings in the earlier studies, the angular momentum of the gas is not negligible, leading to the formation of protostellar disks near the core. As the accretion rates inside these rotationally supported disks ($\dot{M} \lesssim 10^{-2}$ $M_\odot$ yr^{-1}, Bromm 2013) are lower than the gas infalling rates into the halo, the disks become gravitationally unstable, ultimately leading to fragmentation and hence, a multiple stellar system. These new finding opens up the possibility of the formation of binaries or higher-order hierarchical structures (triple, quartet and so on) as well as interactions between those *in-situ* stars. The typical number of the stars found in each halo upon fragmentation is generally less than $6 - 10$ (see Figure 10 in Susa et al. 2014).

As multiple stars can coexist and survive, it now becomes more complicated to determine what the final masses of the stars would be, or the primordial initial mass function (IMF). Determining the IMF for Pop III stars is particularly important because it is closely correlated with their remnant BH masses and their ionizing radiation power (e.g., Eddington luminosity $\propto$ BH mass). Furthermore, the mass of the stars will significantly influence the orbital parameters of stable bound systems as dynamical interactions between the stars are unavoidable for such multi stellar system. If two stars with a comparable mass happen to form in a halo with other less massive stars, the two stars would dominate the dynamics between the stars, likely ejecting all other smaller ones. If the masses of the stars in a halo are comparable, they go through more chaotic interactions.

The initial mass function can be affected by a number of factors, such as different assumptions on feedback regulating accretion rates (Hirano et al., 2014), viscosity in the protostellar disc (Hirano & Bromm, 2017), the first magnetic fields (Machida & Doi, 2013), etc. Overall, it has been suggested that it should be a flat top-heavy function. The IMF is generally

expressed by a power law,

$$\frac{dN}{dM} \sim M^{-\alpha}. \tag{2}$$

Many authors have tried to statistically determine α. Greif et al. (2011) find the mass functions, based on their five independent realizations, to be relatively flat ($\alpha \simeq 0$) within a range of 0.1 $M_\odot$ to 10 $M_\odot$. In Stacy & Bromm 2013, they find a power law fit to their mass distribution with $\alpha \simeq 0.17$, extending from 0.5 $M_\odot$ to 40 $M_\odot$. And Stacy et al. (2016) find a somewhat higher $\alpha \simeq 1 - 1.13$ with a similar mass range as the other studies. Note that $\alpha = 2.35$ for the Salpeter IMF frequently used for present-day stars.

1.1.2 Role of X-rays in the reionization era and HMXBs

The appearance of the first luminous objects marks the end of the cosmic dark ages and the ionizing UV, Lyman α and X-ray photons produced from them lead to the onset of the reionization era. In particular, among those photons, it is expected that X-rays could make a significant contribution to partial reionization and, more importantly, pre-heating of the early intergalactic medium (IGM) (Haiman, 2011; Venkatesan et al., 2001). The thermal history of the IGM can be measured through 21 cm line detection.

X-ray photons have long mean free path. Note that the comoving mean free path of an X-ray photon $\lambda_{\rm X-ray}$ with energy E increases as E^3, i.e., $\lambda_{\rm X-ray} \propto E^3$ (See Equation 116 of Furlanetto 2006). This means that they can escape freely from where they are created and heat a wider spatial range than UV photons. The energy dependence of the mean free path also implies that X-rays behave somewhat differently depending on whether they are soft or hard. Soft X-rays (a few 0.1 keV $< E <$ 2 keV) possibly dominate heating on small scales, whereas hard X-ray photons (2 keV $< E$) propagate further out and influence the global evolution of the thermal history of the IGM. The fraction of the total radiation energy from a source in soft/hard photons clearly depends on its spectral energy density function, which is typically assumed to follow a power low of $E^{-\beta}$ with $\beta = 1$ for a non-thermal component from 'ultraluminous' X-ray sources in nearby galaxies (Kuhlen & Madau, 2005).

Broadly speaking, X-ray heating has two impacts on the subsequent evolution of structure formation (Oh, 2001). On one hand, as the IGM heats up, it is more difficult for gas to collapse to form new stars (or increase the Jeans mass). This can provide negative impact on the formation of dwarf galaxies (Benson & Madau, 2003) and the growth of massive BHs (Tanaka & Haiman, 2009). On the other hand, X-rays can promote cooling by enhancing the formation of H_2, which is the main coolant in this era. Whether the net effect of X-ray heating is negative or positive still remains uncertain. For example, Hummel et al. (2015) study the impact of a cosmic X-ray background on the formation of the first stars. They find that X-ray heating becomes more important at $n < 1$ cm^{-3} whereas the net effect is more likely cooling by H_2 at $n > 10^2$ cm^{-3}. But the overall star formation rate is found to be less sensitive to the strength of the X-ray background. Jeon et al. (2012, 2014), focusing on X-ray emission from accreting BHs, show that the net feedback on star formation is positive, but mild.

One of the important X-ray sources in the local Universe are high-mass X-ray binaries (HMXBs). HMXBs are a class of binaries consisting of a compact object (BH or neutron

star) and a massive companion star (typically > 10 $M_\odot$). They are important objects to understand nucleosynthesis, binary stellar evolution, accretion and energy emission. Generally speaking, from their spatial distribution and metallicity, we can get information about the evolution history of their host galaxies. Given the relatively short evolutionary time scale (~ 3 Myr, the main sequence life time of a vary massive star) of Pop III stars, the X-ray luminosity from HMXBs can be used as a star formation rate (SFR) indicator (Lewin et al., 1997; Shtykovskiy & Gilfanov, 2007). In the local universe, it has been observed that hard X-ray luminosity within a $2-10$ keV band has a linear proportionality with the SFR (Grimm et al., 2003; Mineo et al., 2012) such that,

$$L^{\rm local}_{2-10\ {\rm keV}} \simeq \alpha \times 10^{39} \times \frac{\rm SFR}{\rm M_\odot\ yr^{-1}}\ [\ {\rm erg\ s^{-1}}], \tag{3}$$

where α varies in different observational studies. For example, $\alpha = 2.6$ (Mineo et al., 2012) and 6 (Grimm et al., 2003).

Combining the possibility of multiple Pop III star formation and their generally high masses, it may be natural to consider that Pop III X-ray binaries would be abundant, hence being dominant X-ray sources in the early Universe. Several authors (Jeon et al., 2012; Fragos, Lehmer, Tremmel, et al., 2013; Fragos, Lehmer, Naoz, et al., 2013) have suggested, based on their numerical simulations, that X-ray emission by HMXBs dominate over other possible X-ray sources (e.g. active galactic nuclei and low mass X-ray binary) at $z \gtrsim 2.5-7$. Despite such an important role as X-ray sources in the reionization era, many parameters relevant for HMXBs remain poorly-constrained. One of those parameters is their formation rate. This has been left as a free parameter in many studies (e.g. Xu et al., 2013). Therefore, finding the rate from physically motivated initial conditions is necessary to more accurately gauge the effects of their X-rays. In §3, we infer the formation rate of HMXBs, based on our N-body simulations.

1.2 Formation of intermediate-mass and supermassive black holes - growth from Pop III remnant BHs

Observations of quasars indicate that almost every nearby massive galaxy hosts a supermassive black hole (SMBH) in its center (Kormendy & Ho, 2013). The 14$^{\rm th}$ data release of the Sloan Digital Sky Survey Quasar Catalog (SDSS, Pâris et al., 2017) contains around half million quasars at $0.1 < z < 6.5$, including Sgr A*, the nearest known SMBH believed to be residing in the center of our Galaxy. Thanks to advancement of observational techniques, more distant (high z) quasars have been increasingly found. In the updated PAN-STARRS1 survey data (Bañados et al., 2016), more than 100 quasars in the redshift range $5.6 \lesssim z \lesssim 6.7$ have been registered.

The formation of such massive monsters is a fundamental open question in Astrophysics. The observations of quasars at $z \approx 6-7$ (e.g Fan et al., 2001; Willott et al., 2003, 2010; Mortlock et al., 2011; Venemans et al., 2013; Bañados et al., 2018) indicate that SMBHs with masses of $\sim 10^9$ $M_\odot$ had managed to form in less than a Gyr after the Big Bang. A SMBH with a mass of 8×10^8 $M_\odot$ at $z = 7.5$ was recently reported (Bañados et al., 2018). To explain the origin of SMBHs at the centers of galaxies, a number of possible growth scenarios has

been proposed. They can be broadly divided into three categories: 1) Growth from Pop III star remnants, 2) Collapse from very massive stars with a mass of $\simeq 10^4 - 10^5$ $M_\odot$ ("Direct collapse model") and 3) Dynamical formation in dense stellar clusters.

Among those scenarios, in this dissertation, we focus on the first one (item 1. Growth from Pop III star remnants). We present in §4 the results of our study on the formation of IMBH/SMBH from Pop III star remnants in atomic cooling halos via "hyper-Eddington" accretion.

This scenario 1) assumes the growth of Pop III remnant BHs via accretion and mergers. Considering accretion rates limited to the Eddington accretion rate with a radiative efficienty η, then a seed BH with a mass of $M_{\rm BH,0}$ will grow exponentially over time,

$$M_{\rm BH}(t) = M_{\rm BH,0} e^{\frac{4\pi G}{\eta c \kappa} t} = M_{\rm BH,0} e^{\frac{t}{t_{\rm Sal}}}, \tag{4}$$

where κ is the opacity of the accreting material. Here, $t_{\rm Sal}$ is an e-folding time, also known as the Salpeter time. For a typical value of $\eta = 0.1$, $t_{\rm Sal} \simeq 45$ Myr. Therefore, it takes around 0.7 Gyr for the mass of a seed BH to grow from $M_{\rm BH,0} = 100$ $M_\odot$ to $M_{\rm BH} = 10^9$ $M_\odot$. Hence if near-Eddington accretion continues and the seeds form early in time, it is apparent that there would be a marginally sufficient time to grow the most massive BHs. Such a continuous growth at the Eddington rate, on the other hand, is not always favorable considering other circumstances, such as depletion of gas due to accretion itself or less efficient gas replenishment. Mergers may be able to provide additional mass increase. But growing BHs can be lost due to dynamical ejections or merger kicks. Furthermore, details on the mass of the seed BHs, the duty cycle, the radiative efficiency, negative feedback due to radiation, gas inflow and outflow at galaxy mergers and etc. add more uncertainties to this picture.

One way to circumvent this issue is the growth of a BH at a rate much higher than the Eddington accretion rate (Begelman, 1979; Volonteri & Rees, 2005; Wyithe & Loeb, 2012). For a BH embedded in a dense and optically thick gas medium, as the diffusion time scale of photons becomes longer than the free-fall time of the gas, the outgoing photons are trapped in the inflowing gas flow and advected inward. While the accretion luminosity is limited to the Eddington luminosity, the accretion rate can exceed the Eddington rate as the gas outside where photons are trapped remains unaffected by the radiation from the inside (Begelman, 1979).

Ideally, this "super-Eddington" (or "hyper-Eddington") accretion may be valid under the assumption of spherically symmetric accretion of the gas onto the BH. Many studies have suggested that this is an acceptable assumption for the BHs at the centres of high-redshift galaxies. When the gas arrives near the central region, it ends up with little angular momentum as a result of efficient angular momentum transport, and the cold flow is nearly pressure-supported (Wise et al., 2008). The angular momentum loss can be achieved by several mechanisms, such as gravitational instability (i.e. "bar within bar" instability Shlosman et al., 1989), turbulence (Wise et al., 2008) or merger (Levine et al., 2008). Even if a disc inevitably forms in zero-metallicity halos with a virial temperature $T_{\rm vir} \gtrsim 10^4$ K, it is likely to be fat and dense (Oh & Haiman, 2002). And if the disc scale height is larger than the accretion influence radius (or, "Bondi radius") and the disk has a sufficiently low spin parameter, this can be treated as a plausible mechanism.

As BHs grow in mass and their host halos evolve over time, the assumptions of spherical symmetric accretion and of photon trapping become no longer valid for massive BHs (e.g. $> 10^5$ $M_\odot$ BHs within $\simeq 10^{10}$ $M_\odot$ galaxies, see Figure 3 in Wyithe & Loeb 2012). In other words, it is unlikely that the duration of such rapid growth does not last long enough to grow stellar-mass BHs all the way up to SMBHs at the galactic centers. However, a dramatic increase in mass of seed BHs via super-Eddington accretion to a few $10^2 - 10^3$ $M_\odot$ would allow the subsequent Eddington-limited growth to explain the origin of SMBHs observed at $z \simeq 7$. We will test this scenario in §4, using N-body simulations.

1.3 Formation of supermassive black hole binaries and their mergers

1.3.1 Observation of supermassive binaries and their mergers

Over the last decade, the number of SMBH binary candidates observed at $z < 3$ galaxies has increased (e.g Gitti et al., 2013; Tsalmantza et al., 2011; Tremblay et al., 2016; Bansal et al., 2017)[2]. Among those, sub-parsec SMBH binary candidates have also been identified (e.g Boroson & Lauer, 2009; Liu, Li, & Komossa, 2014; Liu, Shen, et al., 2014; Kharb et al., 2017). Examples are a $10^{7.3} - 10^{8.9}$ $M_\odot$ SMBH binary, separated by 0.1 pc at $z \simeq 3.8$ (Boroson & Lauer, 2009) and a binary black hole with 3.63×10^7 $M_\odot$ separated by 0.35 pc in the Seyfert galaxy NGC 7674 at $z \simeq 0.03$ (Kharb et al., 2017).

Observations of such small binaries, if they are finally confirmed as true SMBH binaries, are of particular interest since whether SMBH binaries can successfully decay to sub-parsec scales remains theoretically uncertain ("final parsec problem", see §1.3.2 below). SMBH binary detections have been made using a number of techniques, including image and photometry analysis from X-ray or radio sources or velocity shifts between narrow and broad emission lines (e.g Decarli et al., 2013). Given limited resolution power, direct imaging searches are restricted to observe somewhat wide binaries in nearby galaxies. Detecting velocity offsets of broad emission lines can be used to identify sub-parsec SMBH binaries (Liu, Li, & Komossa, 2014). However, considering alternative scenarios for velocity displacements, such as radiation from gas in the accretion disk (Strateva et al., 2003) or a recoiled BH due to merger kicks (Bogdanović et al., 2007; Shields et al., 2009), observations of sub-parsec SMBH binaries may not be the smoking gun for the existence of very compact binaries or possible merger events in the near future.

Another important observation in the coming years is detecting gravitational wave signals from mergers of SMBH binaries using the Pulsar Timing Array (PTA). This has been considered as a very promising way to confirm mergers between SMBHs. We discuss this in detail in §5.7.

[2]Some of those reported candidates are found to be separated typically by $\sim$ kpc and assumed not bound, which are also known as "duals" (e.g Chornock et al., 2010; Comerford et al., 2013).

1.3.2 Formation of supermassive black hole to its merger - brief review of dynamics

The journey of a binary from its formation to its coalescence can be divided into three stages. Note that the spatial scale given in the title of each stage is an approximate estimate and it varies depending on details of the binary and its surroundings.

1. *Stage 1. Dynamical friction* ($r \simeq$ a few pc $-$ kpc)

   When two galaxies merge, the black holes which have been residing in the progenitor galaxies spiral into the merged galaxy due to dynamical friction. It is expected that unless the masses of the infalling BHs were small, they would make it to the core regions in less than a Hubble time. One can estimate the time for a BH with a mass $M_{\rm BH}$ to get sufficiently close to the center from a distance r. Assuming a singular isothermal profile for the host galaxy, the infalling time scale $t_{\rm infall}$ can be expressed as follows (Binney & Tremaine, 1987),

$$t_{\rm infall} \simeq 3\ {\rm Gyr} \left(\frac{6}{\ln\Lambda}\right) \left(\frac{r}{5\ {\rm kpc}}\right)^2 \left(\frac{\sigma_{\rm host}}{200\ {\rm km\ s^{-1}}}\right) \left(\frac{M_{\rm BH}}{10^8\ {\rm M_\odot}}\right)^{-1}, \tag{5}$$

   where $\ln\Lambda$ is the Coulomb logarithmic, which measures a spatial range within which the background medium contributes to the frictional force on the BH. And $\sigma_{\rm host}$ refers to the one-dimensional velocity dispersion of the background stars. In a more realistic picture of a merging galaxy, the infalling BH is surrounded by stars when it is sufficiently far from the center. As its orbit decays, the stars bound to the BH will be stripped off due to the tidal forces from the host galaxy. Considering a higher effective mass of the BH and the stars bound to it, the infalling time would become even shorter. When the mutual gravitational potential between the two BHs is dominant over the background potential from the stars, they will become bound to each other and form a binary.

   As Equation 5 implies, the infalling times for less massive BHs can exceed a Hubble time ($t_{\rm infall} \propto M_{\rm BH}^{-1}$). This means that we can expect that a population of wandering satellite BHs outside the core regions exists (Bellovary et al., 2010).

2. *Stage 2. Three-body stellar encounter* ($r \simeq 10^{-2}$ pc - a few pc)

   As the orbit further decays, the surrounding stars start to feel the effect of the binary when the binding energy of the binary becomes comparable to the typical kinetic energy of the stars. Three-body interactions between the binary and each individual star become more important in the subsequent evolution of the binary. The stars take a certain amount of energy from the binary via the slingshot mechanism, resulting in hardening of the binary. However, in this process, surrounding stars are easily ejected. Unless the stars with low angular momentum are replenished faster than they are depleted, the BH binary may experience a lack of stars to interact with. A conical region in parameter space occupied by a population of stars with low angular momentum which can scatter with the binary is called "loss cone". If the loss cone is

not efficiently refilled by e.g., diffusion of stars into the loss cone by 2-body relaxation, the hardening rate of the binary will become constant.

This is likely to be the slowest evolution. Whether or not BHs would be able to coalesce would be mostly determined by this "bottleneck" stage. For a spherically symmetric potential, the time scale for the binary to decay until GW takes over is expected to exceed a Hubble time in many cases (Yu, 2002; Begelman, 1979). The characteristic semi major axis $a_{\rm h}$ of a "hard" binary at this stage is often defined as (Merritt & Wang, 2005),

$$a_{\rm h} = \frac{q}{(1+q)^2}\frac{r_{\rm in}}{4} \simeq 5\ \frac{q}{(1+q)^2}\left(\frac{M_{\rm bin}}{10^8\ {\rm M}_\odot}\right)\left(\frac{\sigma}{200\ {\rm km\ s^{-1}}}\right)^{-2}\ {\rm pc}, \tag{6}$$

where q is the binary mass ratio (≤ 1) and $r_{\rm in}$ is the influence radius at which the enclosed mass of stars around a BH is comparable to the BH mass. Here, σ corresponds to the velocity dispersion of the stars beyond $r_{\rm in}$. The hard radius $a_{\rm h}$ is roughly a $\sim$ pc scale. Thus this is often referred to as the "final parsec problem" (Merritt & Milosavljević, 2005).

3. Stage 3. Gravitational radiation ($r \leq 10^{-2}$ pc)

When the binary separation decays down to a distance at which gravitational wave emission becomes significant, the binary will merge on a short time scale (Peters & Mathews, 1963; Peters, 1964),

$$t_{\rm GW} \simeq 10^{-2}\frac{1}{q(1+q)}\left(\frac{a}{10^{-2}\ {\rm pc}}\right)^4\left(\frac{M_{\rm BH,1}}{10^8\ {\rm M}_\odot}\right)^{-3}(1-e^2)^{7/2}\ {\rm Myr}. \tag{7}$$

When the BHs finally merge, the merged remnant BH gets imparted a recoil kick from anisotropic GW emission and goes through mass loss (Bekenstein, 1973; Fitchett & Detweiler, 1984; Favata et al., 2004). The magnitude of the recoil kick depends on the configuration of the binary, including its spin, and on the orientation of the orbital plane with respect to the spins and on their mass ratios. Recent relativistic simulations found that kick velocities can be larger ($\gtrsim$ 5000 km s^{-1}) than galactic escape velocities due to non-linear spin-orbit coupling, such as partial alignment of the black hole spins with the orbital angular momentum ("hang-up" kick Lousto & Zlochower, 2011; Lousto et al., 2012) or anti-alignment of the spins of the two BHs on the orbital plane ("superkick" Brügmann et al., 2008; González et al., 2007). However, it is still unclear what the final configuration of the spins and the angular momentum will be since spin alignment is subject to torques from the accretion disk (Bogdanović et al., 2007) and to relativistic spin precession (Kesden et al., 2010).

1.3.3 Other mechanisms for mergers: multi-BH interactions and more

Many authors have suggested a number of different hardening mechanisms to bypass the possible bottleneck. For the wet (gas-rich) merger case, it is natural to consider a circumbinary disk around the central binary. It has been shown using hydrodynamical simulations that tidal torques play a role in decaying the orbit of the binary (e.g Mayer et al., 2007; Lodato et al., 2009; Mayer, 2013; Tang et al., 2017). In the presence of the gas disk, the secondary BH migrates on a reasonably short timescale. For example, for a secondary BH massive enough to open a gap in the disk, the BH migrates on a viscous timescale (e.g. 10^7 yr for a binary with masses 5×10^8 $M_\odot$ and 10^7 $M_\odot$ from 0.1 pc to merger in Armitage & Natarajan, 2002).

In a gas-poor environment, it has been suggested that BH binaries can go through more interactions with stars in a non-spherical potential (i.e., axisymmetric or triaxial). In this geometry where the angular momentum of each star is no longer conserved, stars can get arbitrarily close to the central binary during their oscillating motions along the axes of the potential. This results in enhancing hardening rates of the binary (Yu, 2002; Gualandris et al., 2017). Indeed, it has been shown that galaxy merger remnants would maintain triaxiality (Gualandris & Merritt, 2012a) stably for a long time (Poon & Merritt, 2002, 2004). Other than the galaxy geometry, BH mergers can be facilitated by rotation of stellar components. Considering that the vast majority ($\simeq 80\%$) of early-type galaxies are found to be rapidly rotating (Emsellem et al., 2011), understanding the effects of interactions with stars in a rotating potential on the evolution of SMBH binary is important. It appears that depending on whether the host galaxy is counterrotating or cororating with respect to the binary orbit, the eccentricity of the BH binary evolves differently. In a spherical corotating potential, encounters with counterrotating stars tend to increase the eccentricity whereas those with corotating stars are more inclined to circularize the orbit. In general, stars are more likely to be ejected when they are on prograde orbits with the binary. This can be attributed to smaller relative velocities, hence larger cross sections (Iwasawa et al., 2011). Therefore, stars approaching the binary on retrograde orbits and subsequently escaping on prograde orbits will take some angular momentum from the binary more efficiently (Sesana et al., 2011). Holley-Bockelmann & Khan (2015) studied the effects of stellar encounters in a rotating flattened potential and found a similar trend for the eccentricity evolution and additional increases in the hardening rate of SMBH binaries than galaxy geometry alone.

None of those mechanisms above may serve as a solution for the "empty loss cone" problem in a gas-free spherical galaxy. However, one very plausible scenario is multi-body interactions between BHs. If BH binaries fail to merge, one possibly inevitable outcome of continuous galaxy mergers is the formation of a cluster of BHs in the cores. In this case, BH coalescences will be facilitated by two effects: 1) chaotic interactions between the BH binary and other BHs and 2) dramatic increases in the eccentricity by the Kozai mechanism when a hierarchical triple forms (Kozai, 1962). As the central binary goes through encounters with other BHs, its eccentricity tends to grow (Heggie, 1975; Valtonen & Karttunen, 2006; Valtonen & Mikkola, 1991) (such that the values of e^2 become equally likely). The binary becomes more compactly bound by ejections of less massive BHs carrying some energy from the binary. Along with satellite BHs which keep orbiting since the galaxy mergers and BHs recoiled via GW kicks, the ejected BHs via three-body interactions will constitute the population of inter and intragalactic BHs outside the cores. If they form a hierarchical triple, the

Kozai-mechanism causes periodic exchanges between the eccentricity and the inclination of the inner binary. This allows for an initially low e inner binary to loose its energy through GW emission more effectively at every closest approach, ultimately leading to a short merger time (Hoffman & Loeb, 2007). This scenario with more than three BHs in the cores is a very promising scenario in a sense that *this is a natural outcome of the "final parsec" problem and a possible solution for it at the same time*. The observation of a triple candidate in SDSS J150243.09+111557.3 at $z \simeq 0.39$ (Deane et al., 2014) may add more plausibility to this scenario. Using N-body simulations, we investigate this scenario along with a full loss cone case by following galaxy merger trees with a realistic galaxy model in §5. We suggest these multi-SMBH interactions as a "mechanism of last resort" for solving the final parsec problem.

1.4 General overview of multi-body interactions

In stellar dynamics, one important mechanism leading to the formation and hardening of binaries is multi-body interactions, i.e., instantaneous interactions between more than three bodies ($N \geq 3$). Those are also one of the dominant mechanisms for the formation of compactly bound binaries or even BH mergers in the studies presented in this dissertation. Several other studies have investigated few-body interactions ($N \leq 6$), mainly focusing on statistical properties of the end products using numerical scattering experiments with different initial conditions. This approach has proved very powerful and efficient due to enormously increasing computational power.

In general, when two (or more) groups of stars [3] run into each other, depending on their relative impact parameter (or how far their closest distance is), different types of encounters can occur. With a sufficiently small impact parameter, chaotic interactions will ensue. For the encounter between a binary and a single star, depending on the energy changes in the initially bound pair, the interaction can lead to "ionization" (disruption of the binary), "excitation" (softening of the binary), "de-excitation" (hardening of the binary) and "'resonance" (no ejection of stars) (see Figure 7 in Heggie 1975). During the course of interactions, hardening of binaries can be caused by the slingshot mechanism. As less massive stars are ejected, carrying some energy from the remaining sub-stellar system, initially unbound stars can become bound with another star, and bound systems get hardened. If they were chaotically interacting, until stable bound systems and the rest of the unbound in the system are well isolated, binary formation (sometimes binaries with different members after ionization, also called "exchange") and ionization continuously occur. It is likely that soft binaries, defined as binaries whose orbital velocities are less than the typical velocities of surrounding stars, get disrupted by encounters with other stars, whereas hard binaries get hardened further (Heggie, 1975; Hills, 1975). On the other hand, for large impact parameters, the encounters are not accompanied by chaotic interactions, but can lead to binding energy and eccentricity variations [4]. As an example, when a supermassive black hole binary is still wide and it is moving through a background stellar medium, the collective effect of a large number of dis-

[3] I refer to stars here, but the general considerations of this section similarly apply to BH-star and BH-BH interactions.

[4] Heggie & Rasio (1996) show that the change in eccentricity of an initially circular binary due to a distant encounter with a third star on a hyperbolic or parabolic orbit follows a power-law $\delta e \propto (a/r_{\rm p})^{5/2}$ where a is the semimajor axis and $r_{\rm p}$ is the periastron distance of the encounter.

tant (and weak) encounters between each BH and stars, also known as "dynamical friction", operates to decay the binary orbit (i.e., negative energy variations).

Whether or not the eccentricity of a binary increases is determined by relative changes of orbital energy and angular momentum. Generally speaking, if angular momentum loss per orbit is dominant over orbital energy loss, the eccentricity grows while the binary keeps shrinking. This can be understood using an "impulse approximation", i.e., an encounter with a perturber causes an impulsive kick with only a velocity variation Δv, not a position variation, of the subject star. In this approximation, the specific energy and angular momentum changes can be approximated as $\Delta E \simeq v\Delta v$ and $\Delta L \simeq r\Delta v$, respectively. For the same Δv, the angular momentum loss ΔL (the energy loss ΔE) dominates near apocenter (pericenter), resulting in increases (decreases) of e. For example, gravitational wave emission leads to circularization of orbits as the radiation power becomes more enhanced at pericenter.

Considering possibly frequent ionizations and exchanges of members during chaotic interactions, this analytic reasoning based on the impulse approximation may not directly apply for chaotic interactions. It is known that chaotic gravitational encounters tend to excite the eccentricities of outcome binaries. The binary eccentricity distribution formed during three-body interactions follows a thermally-averaged density function (Heggie, 1975),

$$f(e) \simeq 2e, \tag{8}$$

with a median value ~ 0.7 for $0 < e < 1$. This is what is typically found for the binary eccentricity distributions formed during three-body scattering experiments (e.g Mikkola, 1983; Valtonen & Karttunen, 2006). Interestingly, this trend is also found at least for the 2+1+1 outcome during binary-binary scatterings (Ryu et al., 2017a). However, we emphasize caution in concluding from this that a thermal distribution applies to all four-body scatterings. This tendency to produce high eccentricities can make significant differences in the evolution of BH binaries and their gravitational wave emission, which will be shown in §5.

2 Methdology

I developed a N-body code designed to perform standard scattering experiments (see §2.6) and to study more specific astrophysical systems (see §3, 4 and 5). In this section, we describe the numerical scheme implemented in the code and useful algorithm tools.

2.1 Code description

The code can solve the equations of motion in 3-dimensional space, using 4th-order & 5-stage Runge-Kutta-Fehlberg methods (RKF45 method, Erwin 1969) with adaptive time steps. The RKF45 is a very precise and stable integration method among the large class of Runge-Kutta schemes, particularly by adapting the Butcher tableau for Fehlberg's 4(5) method.

To ensure numerical precision, our computational scheme varies the value of each subsequent time step analytically, so that numerical errors for each variable in the simulation do not exceed a certain fraction y of the size of the variable. Typically $y = 10^{-10} - 10^{-13}$. In some cases, however, this method leads to excessive computational effort for calculating relatively trivial interactions. For example, near the pericenter of hyperbolic or highly elliptical encounters, the time steps become increasingly small in a runaway fashion to compensate for the steep rise in acceleration and associated errors. In order to avoid such situations, we implement the following two numerical shortcuts to keep computation times tractable.

The first shortcut is to use analytic approximations for very close 2-body encounters. This is justified in cases where pairs of stars are sufficiently close to each other and isolated from the other stars in the simulation, so that the gravitational pull from the other stars and the background potential are negligible compared to the mutual gravitational pull of the pair. The code employs analytic solutions for any close stellar pairs that satisfy this condition. In cases where the orbital phases are not important, the code can simply replace the binary into one object moving at the center of mass velocity. The code reverts to the RKF algorithm when the conditions stop being satisfied. Different conditions can be added depending on problems of our interest.

Whether two bodies are well isolated from other objects is determined based on a certain fraction, the so called tidal tolerance parameter δ. The relative force $F_{\rm rel}$ and the tidal force $F_{{\rm tid},\star}$ on the $i_{\rm th}$ binary at apocenter are:

$$F_{\rm rel} = \frac{G m_{i1} m_{i2}}{[a(1+e)]^2}\,, \tag{9}$$

$$F_{{\rm tid},\star} = \sum_j \frac{2G m_i m_j}{r_{ij}^3} a(1+e)\,. \tag{10}$$

Hence, the condition for the tidal tolerance parameter can be written:

$$\frac{F_{{\rm tid},\star}}{F_{\rm rel}} < \delta. \tag{11}$$

We generally take $\delta = 10^{-9}$ for the studies presented in this paper. This parameter is also used in termination criteria and binary-tree algorithm, which will be explained later.

Because we are dealing with systems with more than three bodies of different masses, particles frequently form hierarchical triple systems whose motions are affected by the Kozai mechanism (Kozai, 1962). Since our analytic solutions above do not account for such changes in mutual inclination, we limit the use of these solutions to situations where $\tau_{\rm outer} < \tau_{\rm Kozai}$, where $\tau_{\rm outer}$ is the dynamical timescale for the outer pair in the hierarchical triple, and $\tau_{\rm Kozai}$ is the time scale for the Kozai mechanism,

$$\tau_{\rm Kozai} \sim \frac{m_1}{m_2}\frac{P_{1,2}^2}{P_{1,3}}(1-e_{1,2}^2)^{1.5}. \tag{12}$$

The subscript 1 above indicates the primary star of the inner compact binary along with the satellite star denoted by the subscript 3, while the hierarchical tertiary star is indexed by the subscript 2, and P is the orbital period.

Our second shortcut for keeping the simulation runtimes manageable is to set a minimum value for the time step. We choose a physically motivated value, $10^{-6}\times\tau_{\rm dyn,\,min}$, where $\tau_{\rm dyn,\,min}$ is the smallest value of the dynamical time between any two stars in the simulation (that are not being treated by the analytic shortcut above) at a given time step. This procedure is necessary when there are three or more stars interacting at small separations, in which case the analytic approximations above cannot be used.

2.2 Binary tree algorithm

The code can classify all hierarchical structures which exist in the system from unbound systems at each time step. The classification is achieved only based on the relative energies between the objects. The code first identifies the most compactly bound system. The code then iteratively finds a bound system with the next highest relative energy assuming the most compactly bound system as a single object. During this process, the information of all existing bound systems are stored. If there is no bound system, all objects are identified as singles. This binary tree algorithm is not exactly the same as the one implemented in FEWBODY code, specifically designed for scattering experiments (Fregeau et al., 2004), but our tree algorithm follows the same logic of the tree algorithm implemented in FEWBODY code. In addition, the code can calculate the times at which each single star is ejected during chaotic interactions. This is done by evaluating the time-step at which the relative energy between a single particle and the remaining substellar system (binary or triple) becomes positive.

2.3 Dynamical evolution of background potential

The code can handle the dynamical evolution of a background potential as long as the potential is integrable. If a model assumes dissipative mechanisms or energy variations in a background medium such as frictional forces, tidal captures or core collapse, the code implicitly estimates the energy changes due to those mechanisms at each time step, then evolves the potential accordingly. In our study on the formation of SMBH binaries and their mergers, presented in §5, we take into account widening of a background potential due to an energy loss from binaries, also known as "scouring effect". In Figure 24, we show how

the background potential evolves differently with (solid lines) and without (dotted lines) the scouring effect.

2.4 Equation of motion

The code can integrate the equations of motion with various forces and masses of particles which evolve in time. Forces that can be integrated in the code include the following: (i) $\boldsymbol{a}_{\rm N} + \boldsymbol{a}_{\rm PN}$: their mutual gravitational attraction including post-Newtonian terms up to 2.5th order, (ii) $\boldsymbol{a}_{\rm df}$: dynamical friction from the surrounding medium, (iii) $\boldsymbol{a}_{\rm bg}$: the gravitational pull of the background matter and (iv) $\boldsymbol{a}_{\rm mg}$: the deceleration due to BH mass increase with momentum conserved. The resulting equation of motion for the $i^{\rm th}$ BH includes the sum of the five forces:

$$\boldsymbol{a}_i = \boldsymbol{a}_{{\rm N},i} + \boldsymbol{a}_{{\rm PN},i} + \boldsymbol{a}_{{\rm df},i} + \boldsymbol{a}_{{\rm bg},i} + \boldsymbol{a}_{{\rm mg},i} \tag{13}$$

We next describe each contribution in detail.

1. *Mutual gravitational forces between particles*
   We calculate the standard Newtonian gravitational force $\boldsymbol{a}_{\rm N}$ as well as post-Newtonian terms $\boldsymbol{a}_{\rm PN}$,

$$\begin{aligned}\boldsymbol{a}_{\rm gr} &= \boldsymbol{a}_{{\rm N},i} + \boldsymbol{a}_{{\rm PN},i} \\ &= -\sum_{j\neq i} G\ M_{{\rm BH},j}\ \frac{\partial\ \Phi(r_{ij})}{\partial\ r_{ij}}\ \frac{\boldsymbol{r}_i - \mathbf{r}_j}{r_{ij}} \\ &+ \boldsymbol{a}_{{\rm 1PN},i} + \boldsymbol{a}_{{\rm 2PN},i} + \boldsymbol{a}_{{\rm 2.5PN},i},\end{aligned} \tag{14}$$

   where G is the gravitational constant, Φ is the pairwise gravitational potential, $\mathbf{r}_i$ is the displacement of the $i^{\rm th}$ particle from the center, and $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$. In our numerical implementation, we adopt the Plummer softening kernel (Binney & Tremaine, 1987).

   We include post-Newtonian terms $\boldsymbol{a}_{\rm PN}$ up to order 2.5, which account for the loss of orbital energy and angular momentum via gravitational waves, but do not account for spin-orbit or spin-spin coupling. The full expressions for these terms can be found in, e.g., Kupi et al. (2006).

2. *Dynamical friction from background matter*
   When an object moves through a medium, it induces an overdensity of the medium, or wake, behind it. The gravitational force due to the overdense region acts as a dissipative drag on the object's motion. In this study, we consider dynamical friction due to both dark matter (DM) and stars.

   For collisionless medium, we adopt the standard Chandrasekhar formula (Binney & Tremaine, 1987),

$$\boldsymbol{a}_{{\rm df},i} = -4\boldsymbol{\pi}\ \ln\Lambda\ f(X_i)\ \frac{G^2 M_i}{v_i^3}\ \rho(r_i)\ \boldsymbol{v}_i, \tag{15}$$

with

$$f(X_i) \equiv \mathrm{erf}(X_i) - \frac{2}{\sqrt{\pi}}\ X_i\ \exp\left(-X_i^2\right), \tag{16}$$

where $X_i \equiv v_i/(\sqrt{2}\sigma_v)$.

On the other hand, for a collisional medium (i.e. gas medium), dynamical friction is generally smaller than that for a collisionless medium whereas it becomes enhanced as a result of resonance of a body moving at a transonic or supersonic speed with respect to the medium. It is because for the collisional medium, sound waves can propagate both downwind and upwind, resulting in a symmetric distribution around the moving body. This ultimately decrease the efficiency of the gravitational drag. On the other hand, for $\mathcal{M} \gtrsim 1$, the drag becomes stronger as the downwind sound waves resonantly interact with the perturber. To account for this, we adopt the modified formula for collisional medium derived by (Tanaka & Haiman, 2009). This prescription incorporates behaviors found in numerical simulations for subsonic and supersonic regimes (Ostriker, 1999a; Escala et al., 2004). The specific drag force vector always points opposite to the direction of motion, and is given by:

$$a_{\mathrm{df},i}^{(\mathrm{coll.})} = -4\pi\ G^2\ M_i\ \rho(\vec{r}_i)\ \frac{1}{v_i^2} \times f^{(\mathrm{coll.})}(\mathcal{M}_i), \tag{17}$$

with

$$f^{(\mathrm{coll.})}(\mathcal{M}_i) = \begin{cases} 0.5\ \ln\Lambda\left[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i\ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\right] \\ \qquad\qquad 0 \leqslant \mathcal{M}_i \leqslant 0.8; \\ 1.5\ \ln\Lambda\left[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i\ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\right] \\ \qquad\qquad 0.8 \leqslant \mathcal{M} \leqslant \mathcal{M}_{\mathrm{eq}}; \\ \frac{1}{2}\ln\left(1 - \frac{1}{\mathcal{M}_i^2}\right) + \ln\Lambda \\ \qquad\qquad\qquad\qquad \mathcal{M}_i > \mathcal{M}_{\mathrm{eq}}. \end{cases} \tag{18}$$

Above, $\mathcal{M}_i \equiv v_i/c_{\mathrm{s}}$ is the Mach number, and c_{s} is the sound speed. $\mathcal{M}_{\mathrm{eq}}$ is determined such that the second and the third equations are continuous for a given choice of $\ln\lambda$.

Note that the expressions for dynamical friction given above (collisional + collisionless) are based on the assumption of non-accelerated motion in a medium with a uniform density distribution. In a system with BH/stars chaotically interacting and forming bound systems, this assumption may not be justified. However, a more accurate treatment of dynamical friction which takes into account the nonlinear evolution of the overdense region and all the acting forces at each time step is beyond what a few-body code is able to handle. On the other hand, since several studies of dynamical friction in a nonuniform medium or for perturbers on nonlinear trajectories (e.g Sánchez-Salcedo & Brandenburg, 2001; Just & Peñarrubia, 2005; H. Kim & Kim, 2007; W.-T. Kim, 2010) show that those trajectories are well described by the Chandrasekhar formula, we will use the formulas above.

3. *Gravitational force of the background matter*
   The background media exert an additional gravitational force on the particles. For a spherically symmetric density profile, this force points toward the centre of the potential. It can be expressed as

$$\boldsymbol{a}_{\mathrm{bg},i} = -\frac{G\ M_{\mathrm{en},i}(r \leq r_i)}{r_i^3}\mathbf{r}_i, \tag{19}$$

   where $\mathbf{r}_i$ is a vector pointing from the centre to the i-th particle and $M_{\mathrm{en}}(r \leq r_i)$ is the enclosed mass inside $r = r_i$..

4. *Deceleration due to mass growth*
   When the masses of particles increase, we take into account the decrease in velocity accordingly. Assuming particles grow in mass in a spherically symmetric fashion, the $i-$th particle decelerates through conservation of linear momentum,

$$\boldsymbol{a}_{\mathrm{mg},i} = -\frac{M_i' - M_i}{M_i'\Delta t}\boldsymbol{v}_i\,, \tag{20}$$

   where M_i' is the increased mass estimated and Δt is the time step.

Each of the forces above can be freely included or excluded in integrating the equations of motion according to specific models of interest. Given the solutions of the equation of motion at every time step, we update the positions and velocities for each particle.

2.5 Code tests

2.5.1 Comparison with numerical results from another code - Outcome fraction

We performed systematic comparative checks between our code and the publicly available N-body code "FEWBODY"[5] during our first study of the background potential (Ryu et al., 2017a), and we confirmed that both codes give statistically consistent results. Here we are presenting a further test. In this test, using our code, we first performed numerical experiments of encounters between two identical binaries ($a = 1$ AU, $e = 0$ and each stellar mass=1 $\mathrm{M}_\odot$) in a homogeneous background potential for various number densities (hence same background potential model described in described in 2.6.1). Taking the same initial conditions as used in Leigh et al. (2016), we explored the nearly-zero total energy range by changing the density (slightly positive total energy at high density to slightly negative total energy at low density). As a consistency check, we ran simulations without the background potential (i.e. number density $n = 0$) and reproduced the results from Leigh et al. (2016), especially the outcome fractions at $E \simeq 0$ in Figure 4 of their paper.

Figure 1 shows the outcome fractions of $2 + 2$ encounters with/without the background potential. As the total energy goes from positive to negative (zero at $n = 10^6$ cm^{-3}), the fractions of the $1 + 1 + 1 + 1$ outcomes drop to zero and the outcome fractions converge to

[5] http://www.ascl.net/1208.011

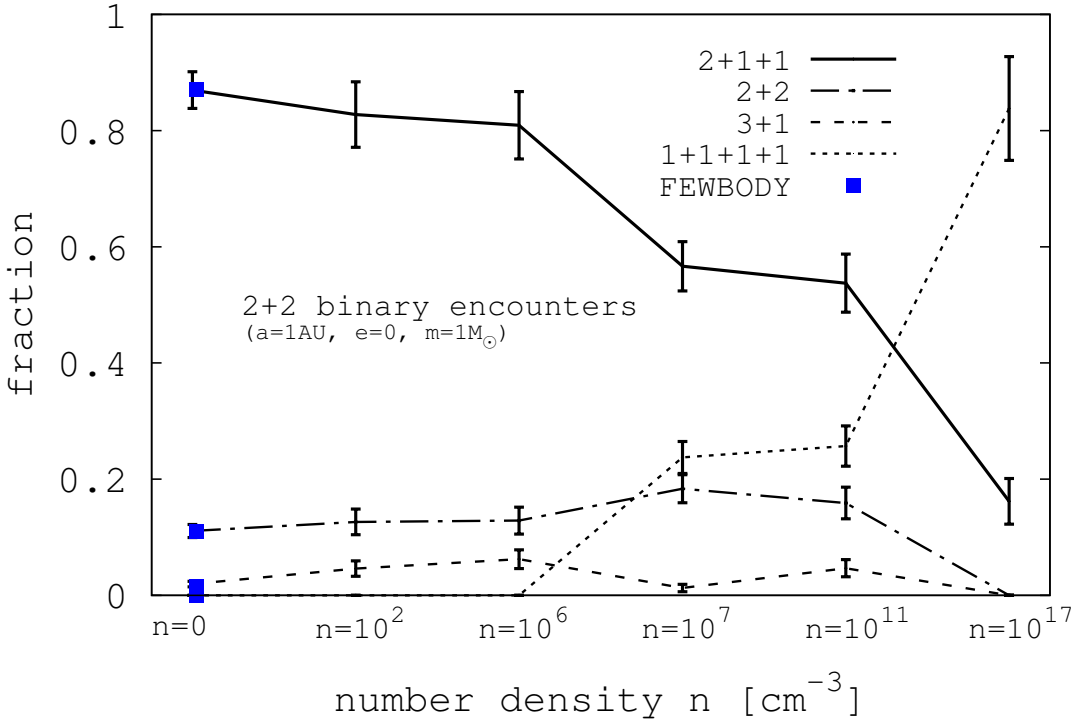

Figure 1: The outcome fractions of $2+2$ encounters with/without the background potential. The total energy is zero at $n = 10^6$ cm^{-3} and becomes positive at $n > 10^6$ cm^{-3} to negative at $n < 10^6$ cm^{-3}. As the total energy becomes negative (smaller n), the fractions of the $1+1+1+1$ outcomes drop to zero and all the outcome fractions converge to those from Leigh et al. (2016) at $n = 0$ (blue squares denoted by "FEWBODY", $2+1+1$, $2+2$, $3+1$ from top to bottom) for the same total energy.

those from Leigh et al. (2016) at $n = 0$ (blue squares denoted by "FEWBODY", $2+1+1$, $2+2$, $3+1$ from top to bottom) for the same total energy.

2.5.2 Comparison with analytic solution

As a further test of our code, and to gain a better physical insight, we next compare the results from a numerical binary-binary scattering with an analytic solution for the distributions of the outcome products. In Ryu et al. (2017d), we derived an analytic formula relating the angle between the velocity vectors of the two ejected single stars and the orbital separation of the remaining binary, both in isolation and in a homogeneous background potential. Two single stars and one binary is a generally frequent outcome of the four possible outcomes when two binaries scatter. In deriving the formula for the outcomes in the presence a background potential, several approximations are made. However, for interactions without the background potential, one can derive an exact solution for the momentum p's of all of the outcomes and their relative angles assuming momentum conservation,

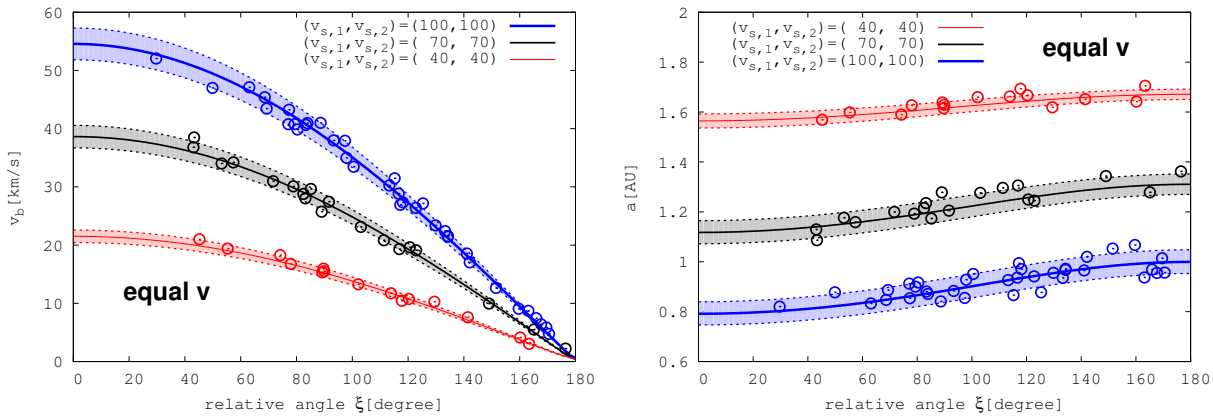

Figure 2: The speeds of the final binaries $v_{\rm b}$ (*left* panel), their semimajor axes a (*right* panel) expected from our analytic formulae 21 and 23 (solid/thick dotted lines) for equal values of the initial velocities ($v_{\rm s,1}$, $v_{\rm s,2}$) in units of km s^{-1}. The data points (dots) show the results of our numerical binary-binary scattering simulations performed using the code. For the simulated data points, the final speeds of the ejected stars are constrained to be within $\sim 5\%$ of the speeds shown in the plots. The solid lines for $v_{\rm b}$ and a are estimated using the median values of the speeds of the ejected stars. For each combination of speeds, we show the upper and lower limits (thin dotted lines) around the median values (solid lines and thick dotted lines), adopting the same colors as for the dots. Furthermore, we fill the regions between the two limits with the same color.

$$p_{\rm s,2} = \sqrt{p_{\rm s,1}^2 + p_{\rm b}^2 + 2p_{\rm s,1}p_{\rm b}\cos\Psi}, \tag{21}$$

$$\cos\Psi = -\frac{p_{\rm s,1} + p_{\rm s,2}\cos\xi}{\sqrt{p_{\rm s,1}^2 + p_{\rm s,2}^2 + 2p_{\rm s,1}p_{\rm s,2}\cos\xi}}, \tag{22}$$

and the semimajor axis a of the binary assuming conservation of the total energy defined as follows,

$$\begin{aligned} E_{\rm tot} &= KE_{\rm s,1} + KE_{\rm s,2} + KE_{\rm b} + E_{\rm bind} \\ &= \frac{1}{2}m_{\rm s,1}v_{\rm s,1}^2 + \frac{1}{2}m_{\rm s,2}v_{\rm s,2}^2 + \frac{1}{2}m_{\rm b}v_{\rm b}^2 - \frac{Gm_{\rm b,1}m_{\rm b,2}}{2a}, \end{aligned} \tag{23}$$

where the subscripts $s,1$ and $s,2$ indicate the two ejected singles and b the remaining binary. ξ is the relative angle between the two ejected single stars and Ψ the relative angle between one ejected star and the binary. The angle between the other ejected single star and the binary (denoted by Φ) is automatically determined since $\xi + \Psi + \Phi = 2\pi$.

In Figure 2, we show the speeds of the final binaries $v_{\rm b}$ (*left* panel), the corresponding semimajor axes a (*right* panel), given by our analytic formulae (Equations 21 and 23, solid/dotted lines) for equal values of the ejected single star speeds ($v_{\rm s,1}$, $v_{\rm s,2}$) in unit of km s^{-1}. To compare, we mark via the data points the results of the numerical experiments

of binary-binary scatterings (2+1+1 outcome) performed using the code. For the simulated data points, the final speeds of the ejected stars are constrained to be within 5% of the speeds shown in the plots. The lines for $v_{\rm b}$, a are estimated using the median values of the speeds of the ejected stars. Hence the actual combinations of the speeds for the estimates are slightly different from those indicated in the plots. For the analytic calculations, we assume the same total initial energy and masses for the stars as adopted in the simulated data. For each combination of the single star ejection speeds, the lines and dots share the same color. All of the parameters $v_{\rm b}$ and a calculated from our analytic formulae show excellent agreement with the numerical simulations. See Ryu et al. (2017d) for more comparisons for various combinations of $v_{\rm b}$, a and Φ.

2.6 Applications of the code – binary-binary scattering experiment

2.6.1 Scattering experiments in a background potential

T. Ryu, N. Leigh, R. Perna *The Monthly Notices of the Royal Astronomical Society*, Vol. 467, 4447 (2017)

Using the N-body code presented above, we perform scattering experiments between two binaries in a background potential. Our primary goal of these experiments is characterize the effects of a continuous background potential on the time evolution of binary-binary interactions. To this end, we present the results of a large suite of numerical N-body simulations of binary-binary scatterings in a background potential. For simplicity, we model the background potential as an uniform (natal or star-forming) gas medium.

We perform 10^3 scattering experiments of two identical binaries with the same initial semimajor axis for every given set of simulations. The number of simulations for each set is sufficiently large to ensure that all data statistically converge (the overall Poisson uncertainties $\lesssim 3-4\%$). In total, we consider three discrete values of the semimajor axis ($a_0 = 1$ AU, 10 AU and 100 AU). We first generate two identical binaries with initially circular orbits (the eccentricities are zero, or $e = 0$). Each binary consists of two point particles, each with a mass of 1 $M_\odot$. Then we give an initial separation and relative velocity between the centre of mass of the two binaries, such that they collide head-on at the origin (i.e., with impact parameter $b = 0$) with nearly zero total energy (more precisely, slightly negative to positive depending on the number density of the gas medium). However, the mutual inclinations between the binary orbital planes, as well as their initial phases, are randomly chosen.

In order to choose the value of the initial relative velocity $v_{\rm rel}$ between the centres of mass of the two binaries, we define the critical velocity $v_{\rm cri}$ as the relative velocity between the centres of mass of the two binaries at infinity for which the total encounter energy is zero. In this paper, $v_{\rm rel}$ will be presented in units of $v_{\rm cri}$, unless otherwise stated. More explicitly, in the case of binary-binary encounters, the total initial energy E_i at infinity is written as

follows:

$$E_i = \frac{1}{2}\mu v_{\rm rel}^2 - \frac{Gm_{11}m_{12}}{2a_1} - \frac{Gm_{21}m_{22}}{2a_2}. \tag{24}$$

In this equation, the first subscript for the particles mass m, along with the binary semimajor axis a subscript, corresponds to each binary, whereas the second subscript for m corresponds to each component of the binary. For example, the masses of the two stars in the $1_{\rm st}$ binary with semimajor axis a_1 are m_{11} and m_{12}. Here, μ in the first term on the right hand side is the reduced mass of the two binaries, i.e., $\mu = (m_{11}+m_{12})(m_{21}+m_{22})/(m_{11}+m_{12}+m_{21}+m_{22}) = 1\ {\rm M}_\odot$. The second and third terms are the orbital energies of the two binaries. Note that the gravitational potential between the two binaries is zero at infinity. Therefore, the critical velocity at $E_i = 0$ is,

$$v_{\rm cri} = v_{\rm rel}(E_i = 0) = \sqrt{\frac{2}{\mu}\Big(\frac{Gm_{11}m_{12}}{2a_1} + \frac{Gm_{21}m_{22}}{2a_2}\Big)}\,. \tag{25}$$

In this study, we take $v_{\rm rel} = 1$ to explore the zero-energy limit.

In the presence of a background gas medium, the condition $v_{\rm rel} = 1$ does not necessarily imply that the total encounter energy is exactly zero. Instead, we achieve the zero-energy limit by adjusting the initial separations between the binary centers of mass. In order to explain how we determine the initial distance between the two binaries, we first need to introduce the background potential adopted in this study, corresponding to a constant density gas medium.

We assume a uniform (i.e. constant) density ρ (n is the gas number density) for this potential, with a limiting outer boundary $r_{\rm bg}$ set by our chosen total background mass $M_{\rm gas}$ such that

$$\rho = \begin{cases} nm_{\rm H} & r \le r_{\rm bg}\,; \\ 0 & r > r_{\rm bg}\,, \end{cases} \tag{26}$$

where $m_{\rm H}$ is the mass of a hydrogen atom. For simplicity, we take a mean molecular weight of unity. The mass of gas enclosed in a spherical volume of radius r can be written as

$$M_{\rm en,gas}(r) = \begin{cases} \frac{4\pi}{3}\rho r^3 & r \le r_{\rm bg}\,; \\ \frac{4\pi}{3}\rho r_{\rm bg}^3 = M_{\rm gas} & r > r_{\rm bg}\,. \end{cases} \tag{27}$$

We consider a range of densities, motivated by the values typical of gaseous astrophysical environments, most notably giant molecular clouds and star-forming regions. In this case, the gravitational force imparted by a uniform gas medium of constant density on a given

star particle at r follows the analytic formula:

$$\begin{aligned}\mathbf{f}_{\rm bg}(r) &= -\frac{GmM_{\rm en,gas}(r)}{r^3}\mathbf{r} \\ &= \begin{cases} -\frac{4}{3}\pi Gm\rho\ \mathbf{r} & r \le r_{\rm bg}\,; \\ -\frac{4}{3}\pi Gm\rho\left(\frac{r_{\rm bg}}{r}\right)^3\mathbf{r} & r > r_{\rm bg}\,, \end{cases}\end{aligned} \tag{28}$$

where m is the mass of the star and $\mathbf{r}$ is the vector pointing from the system center of mass (CM) to the star. Accordingly, the background potential has the following form:

$$V_{\rm bg}(r) = \begin{cases} \frac{2}{3}\pi Gm\rho(r^2 - 3r_{\rm bg}^2) & r \le r_{\rm bg}\,; \\ -\frac{GmM_{\rm gas}}{r} = -\frac{4}{3}\pi Gm\rho\frac{r_{\rm bg}^3}{r} & r > r_{\rm bg}\,. \end{cases} \tag{29}$$

And the total energy $E(t)$ of four stars in the system at t, including the contribution from $V_{\rm bg}(r)$, can be written as,

$$E(t) = \sum_{i=1}^{4}\frac{1}{2}m_i v_i^2 - \sum_{\substack{i,j=1 \\ (i>j)}}^{4}\frac{Gm_i m_j}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{4}V_{\rm bg}(r_i)\,, \tag{30}$$

where m_i is the mass of each star and v_i is the velocity of each star with respect to the system CM.

In order to determine the initial separations r_{12} between the centres of mass of the two binaries (located at initial distances r_1 and r_2 from the system CM), we use the monopole approximation for the binaries (i.e. a point particle with a mass of 2 $M_\odot$) in equation 30 to evaluate both of the gravitational potential energies (between the two binaries and from the background medium). Under the presence of a background potential, this approximation allows us to derive r_{12} by solving one simple equation for a given total energy. More explicitly, the initial total energy for the two binaries in the background potential becomes:

$$\begin{aligned} E_0 &= \frac{1}{2}\mu v_{\rm rel}^2 - \frac{Gm_{11}m_{12}}{2a_1} - \frac{Gm_{21}m_{22}}{2a_2} \\ &\quad + \frac{2}{3}\pi Gm_1\rho r_1^2 + \frac{2}{3}\pi Gm_2\rho r_2^2 - \frac{Gm_1m_2}{r_{12}} \\ &= \frac{1}{2}\mu v_{\rm rel}^2 - 2\frac{Gm_{11}m_{12}}{2a_1} + \frac{4}{3}\pi Gm_1\rho\left(\frac{r_{12}}{2}\right)^2 - \frac{Gm_1m_2}{r_{12}}\,. \end{aligned} \tag{31}$$

In the second equality, we simplified the equation using the fact that the two binaries are identical and the CM of the two binaries coincides with the origin initially, i.e. $r_1 = r_2 = r_{12}/2$. Here, the gravitational potential energy between the two binaries (the last term) should not be ignored because the two binaries are initially separated by a finite distance in our simulations. Since we take $v_{\rm rel} = 1$, the first two terms on the right hand-side of the second equality (corresponding to E_i in equation 24) always cancel out, regardless of a_1 (and $a_2 = a_1$). Hence, in the limit of $E_0 \simeq 0$, r_{12} is determined by the requirement that the last

two terms cancel each other. Once r_{12} and the relative position of each star with respect to the center of mass of each binary are decided, so are the position and velocity vectors of the four stars.

In our experiments, we fix r_{12} such that $E_0 \simeq 0$ (but slightly positive) when $n = 10^2$ cm^{-3}. In order to explore the effects of the background potential, we run suites of simulations with a range of densities, i.e., $n = 10, 10^2, 10^3$ and 10^5 cm^{-3} (default set). For this default set, we take $M_{\rm gas} = 10^{10}$ M$_\odot$ so that the width of the potential ($r_{\rm bg}$) is large enough for all particles to remain within the spherical volume defined by this critical radius throughout the duration of the interaction. For comparison, we run additional simulations with even higher number densities and larger semimajor axes. As we will show, the background potential resets the zero-point of the total system energy. In particular, the stellar dynamics are significantly affected by the background potential when the stars become trapped in the potential, oscillating with a short period around the system CM. This leads to repeated scattering events that would not otherwise occur without the influence of the background potential.

The simulations terminate when no further scatterings will occur. To begin, we use the same criteria as described in Fregeau et al. (2004). These criteria consist of four conditions: 1) All single stars and bound hierarchies in a system are moving away from one another. 2) The total energy of the single stars and bound hierarchies are positive (without including the orbital energies of any bound hierarchies). 3) Hierarchies such as triples are dynamically stable. Finally, 4) the relative force between stars in a bound system should be larger at apocenter than the tidal force from all other stars (i.e., bound hierarchies as a whole, if formed) by a tidal tolerance parameter δ. We take $\delta = 10^{-9}$, which is a value small enough to allow us to precisely classify the outcomes and explore the effects of the background potential at large distances. For more details, see Fregeau et al. (2004). 5) $v \geq v_{\rm esc} = \sqrt{\frac{2GM_{\rm en}(r)}{r}}$ if $r \geq r_{\rm bg}$,
where $M_{\rm en}(r)$ is the total mass of gas and stars enclosed in a spherical volume of radius r and $v_{\rm esc}$ is the velocity required by a star at r to escape from the gas medium to spatial infinity.

We classify the final end products of binary-binary scatterings according to the four outcomes:

1. a binary and two single stars (2+1+1)
2. two binaries (2+2)
3. a triple and a single star (3+1)
4. four single stars (1+1+1+1)

Henceforth, we will refer to each outcome by what is given in the accompanying parentheses.

Figure 3 shows each outcome probability as a function of gas density for each semimajor axis. The error bars indicate the Poisson uncertainties for each simulation set. The line types differentiate between the different semimajor axes: the solid, dot-dashed and dotted lines correspond to, respectively, $a_0 = 1$ AU, 10 AU and 100 AU. The different point types indicate the different outcomes: the triangles, circles, down-pointing triangles and squares correspond to, respectively, 2+1+1, 2+2, 3+1 and 1+1+1+1.

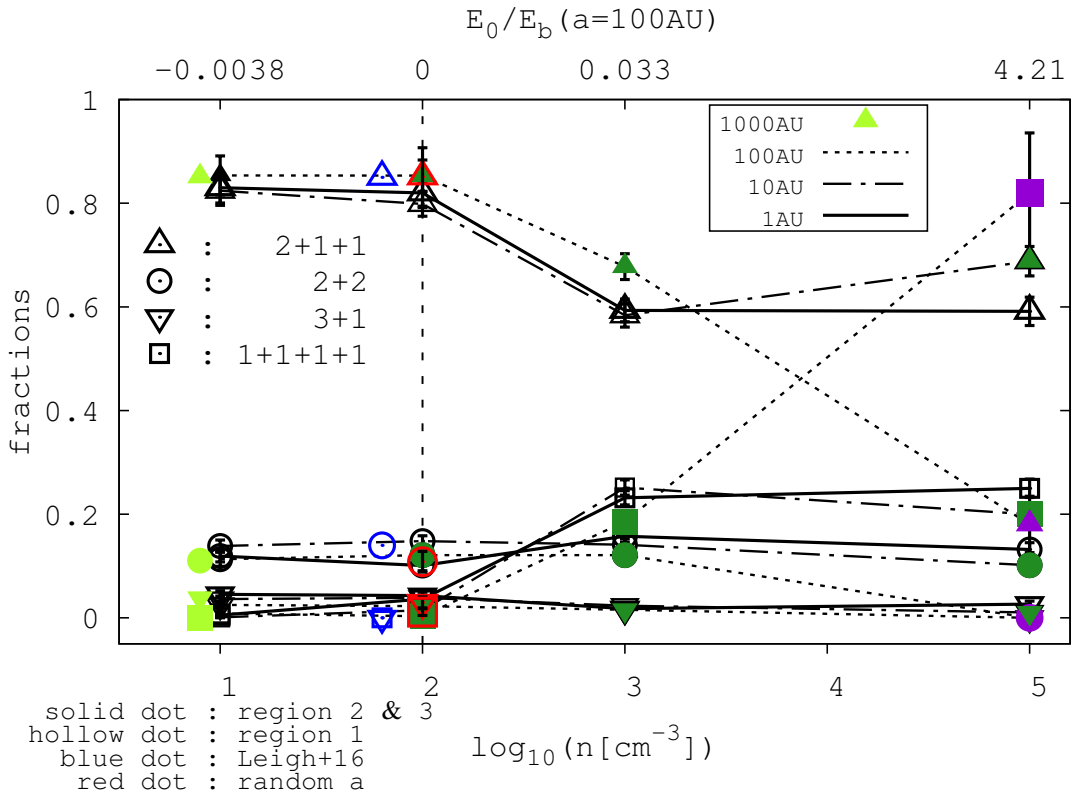

Figure 3: The fraction of each outcome as a function of gas density for three discrete values of the semimajor axs. The different line types correspond to the different semimajor axes: the solid line, dot-dashed line and dotted line represent, respectively, $a_0 = 1$ AU, 10 AU and 100 AU. The different point types indicate the different encounter outcomes: the triangles, circles, down-pointing triangles and square dots correspond to, respetively, the outcomes 2+1+1, 2+2, 3+1 and 1+1+1+1. The outcome fractions at $n = 0$ cm^{-3} and $E_0 = 0$ in Leigh et al. (2016) (see Figure 4 in their paper assuming a virial ratio $k \simeq 1$) are marked as blue dots, and those for our simulations where the semimajor axis is randomly generated in the range of $a_0 = 1 - 100$ AU are marked as red dots. The bright green dots near $n = 10$ cm^{-3} represent the fractions for an additional experiment with $n = 10$ cm^{-3} and $a_0 = 1000$ AU. The setup details are described in the text.

1. *Exploring the $E_0 = 0$ neighborhood*

   Figure 3 shows that the outcome fractions of our numerical study at $n = 10 - 10^2$ cm^{-3} are almost the same as for the equivalent scattering experiments without a background potential (i.e., with $n = 0$). As n increases from $n = 10^2$ cm^{-3} to $n = 10^3$ cm^{-3}, the total energy becomes positive and the 1+1+1+1 outcome begins to appear for all semimajor axes. To ensure that we have correctly calculated the outcome fractions at this critical energy limit, we run two more sets of simulations: 1) Instead of assuming discrete values for the semimajor axes, we generate two binaries with the same semimajor axis, but randomly generated within the range $a_0 = 1$ AU $- 100$ AU. The results of this experiment (marked with red hollow dots at $n = 10^2$ cm^{-3} in the figure) are consistent with those found adopting discrete values for the semimajor axis; 2) we performed experiments with a density between the two critical densities, or $n \simeq 300$ cm^{-3} ($\log[n/\text{ cm}^{-3}] = 2.5$) for both $a_0 = 1$ AU and 10 AU. As expected, the outcome fractions for $n \simeq 300$ cm^{-3} fall between those for $n = 10^2$ cm^{-3} and $n = 10^3$ cm^{-3} in all cases. We do not show them to avoid overcrowding the figure. Therefore, both experiments show that the outcome fractions and their dependences on the gas density are consistent with the ranges and values physically allowed by the total energy (e.g., the 1+1+1+1 outcome only occurs for positive total energies).

   We compare our results with the outcome probabilities at $n = 0$ cm^{-3} and $E_0 = 0$ given in Leigh et al. (2016) (see Figure 4 at virial ratio $k \simeq 1$). These are marked as blue hollow dots using the same shape for each outcome. To avoid overlapping with other data points, these points are marked off the $E_0 = 0$ line (the vertical dotted line at $n = 10^2$ cm^{-3}). Note that the numerical scattering experiments in Leigh et al. (2016) agree with those shown here for $a_0 = 1$ AU since all assumptions are the same, including the semimajor axis ($a_0 = 1$ AU), the eccentricity ($e = 0$), the stellar mass ($M = 1$ M$_\odot$) and the total initial energy ($E = 0$). Additionally, we find that the outcome fractions in our study at $n = 10 - 10^2$ cm^{-3} are still comparable to those found in Leigh et al. (2016), even for our other choices of the semimajor axis ($a_0 = 10$ AU and 100 AU). It is because the gravitational forces between stars dominate, and hence the stellar dynamics should remain the same as in the absence of the background potential.

   However, as the density increases, the final encounter outcomes and their probabilities change, due to the effect of the background potential, which acts to reset the zero-point of the total system energy. The fractions, especially for $a_0 = 100$ AU, diverge slightly from the results for the other semimajor axes as n increases. We start to see that the ejected stars return toward the system CM. Consequently, the stars take part in repeated scattering events that would not have occurred in the absence of a background potential. This leads to discrepancies in the final outcome probabilities relative to the no background potential case. This regime is different from the case where stellar interactions dominate.

   Interestingly, just above the zero-energy limit, the fraction of 2+1+1 outcomes decreases by about the same amount that the 1+1+1+1 outcome increases, whereas the fractions of the other two outcomes (3+1 and 2+2) remain the same. The probability of having a 1+1+1+1 outcome appears to be determined by whether or not all four

stars are promptly ejected. In other words, if the stars are ejected one at a time, then a 1+1+1+1 outcome does not occur. This is because once a single star is ejected with positive kinetic energy, then the left-over system should have a negative total energy, which can never end up fully ionized (by energy conservation). Note, however, that if the initial total energy is positive and sufficiently large, it is possible that after an ejection event the remaining subsystem still has a positive total energy, and so can be reionized into single stars later on. If $n = 10^3$ cm^{-3}, then the total energy is much smaller than the typical kinetic energies of ejected single stars (estimated as $v_{\rm esc} \sim v_{\rm cri}$). Thus, we expect that the prompt ejection of all four stars is the only channel for full ionization. As the 1+1+1+1 outcome begins to appear, the fractions of all other outcomes decrease (by what appears to be a roughly constant fraction). Since the 2+1+1 outcome probability is the highest, its decrease looks especially significant.

One other point to note in the density range $n = 10$ cm^{-3} to $n = 10^3$ cm^{-3} is that the increase we find in the probability of having a 1+1+1+1 outcome (i.e., from 0 to 0.2) is relatively large compared to the results of other scattering experiments (Mikkola, 1983; Fregeau et al., 2004). This discrepancy is most likely due to the presence of the background potential, but the different choices for the impact parameter and eccentricity could also contribute. In our study, we only consider initially circular orbits and zero impact parameters (i.e., head-on collisions). However, other studies draw the eccentricity and impact parameter from specified distributions. For instance, when two binaries collide head-on there is a higher probability of full ionization occurring relative to encounters with non-zero impact parameter, since all four stars are more likely to interact strongly at the first encounter such that the two binaries would both be (at least temporarily) fully ionized. Therefore, initial configurations such as head-on collisions can increase the fraction of 1+1+1+1 outcomes. Meanwhile, the background potential serves to enhance the effects of gravitational focusing, drawing all stars toward the system CM.

2. Exploring the strong-background potential, positive-energy regime

   The outcome probabilities become different for each semimajor axis. This is different from the cases with $n = 10 - 10^3$ cm^{-3} in which each outcome probability for all the semimajor axes consistent within one or two standard deviations as long as n is same. More explicitly, for $a_0 = 1$ AU, these probabilities are almost the same as for $n = 10^3$ cm^{-3}. Since for this value of a_0 the gravitational forces between the stars still dominate over the background potential. the outcome probabilities are expected to remain the same as those for $n = 0$. Since the cross section for each outcome does not vary much Sweatman (2007) from $n = 10^3$ cm^{-3} to $n = 10^5$ cm^{-3} (0 - 0.04 in the unit of $E_{\rm b}(a = 1$ AU)), these outcome probabilities are reasonable. Notice that the units shown along the upper x-axis of the figure correspond to $E_{\rm b}(a = 100$ AU). For $a_0 = 10$ AU, the outcome probabilities start to differ, similar to the cases for $n = 10^2$ and 10^3 cm^{-3} cases with $a_0 = 100$ AU.

   Interestingly, for $a_0 = 100$ AU, the outcome probabilities for 2+1+1 and 1+1+1+1 reverse and all other outcomes (3+1 and 2+2) are completely suppressed. For the cases with the 2+1+1 outcome, we find that subsequent stellar scatterings continue,

but more frequently due to the stronger potential and hence higher accelerations, until two single stars are completely ejected from the system. Once a compact binary forms with sufficiently large binding energy, it survives successive encounters, becoming more and more compact via the slingshot mechanism. This also implies that it is more likely that a wide binary or triple will be ionized (i.e., the complete suppression of the 2+2 and 3+1 outcomes). In the presence of a deep potential, a single star has to gain a higher escape velocity for complete escape, necessarily leading to a hardened binary. Recall that $v_{\rm esc} \sim \sqrt{\rho}$ (see the additional termination criteria 6). As a result, we see a shift of the binding energy distribution towards the higher energy range. Otherwise, like the cases with the 1+1+1+1 outcome, two binaries should be first ionized into four stars in order to escape.

3. Additional experiments

   We performed one additional experiment to more clearly understand the effects of the background potential on the subsequent stellar dynamics for negative total energies. In this experiment, we assume $a_0 = 100$ AU and $n = 10^7$ cm^{-3}.

   For this particular experiment, we set $M_{\rm gas} = 40$ M$_\odot$($r_{\rm bg} = 6600$ AU). This choice for the total gas mass is such that stars can escape from the background potential well and the effects of the background potential should be significant. In this experiment the only end product is 2+1+1 (cf. 2+1+1 and 1+1+1+1 in the simulation for $E_0 > 0$ with $n = 10^5$ cm^{-3} and $a_0 = 100$ AU). It is possible that two binaries are fully ionized but given the total energy budget available to all four single stars, they cannot have sufficiently high escape velocities (or kinetic energies) to meet the criterion for all four single stars to pass the outer potential boundary.

In summary, we find that one of the important roles of the background potential (for any non-zero number density) is to reset the zero-point of the total system energy. This in turn affects the types of objects formed during the encounters in addition to the outcome probabilities. In particular, for higher n and less tightly bound systems (i.e., larger semimajor axes), stars often become trapped in the background potential, oscillating with a short period around the system center of mass. This results in fewer triples and wider binaries and the preferential survival of compact binaries, since subsequent scatterings continue until two single stars escape at very high ejection velocities.

2.6.2 Formation of runaway stars

T. Ryu, N. Leigh, R. Perna *The Monthly Notices of the Royal Astronomical Society*, Vol. 470, 2049 (2017)
T. Ryu, N. Leigh, R. Perna *The Monthly Notices of the Royal Astronomical Society*, Vol. 470, 2 (2017)

Runaway stars are a population of fast-moving stars, characterized by speeds $\gtrsim 30$ km s^{-1}. They are generally of the O and B spectral type, and are often found at some distance from

Mass (binary 1, binary 2)	$[39\ \mathrm{M}_\odot + 16\ \mathrm{M}_\odot]$, $[19\ \mathrm{M}_\odot + 16\ \mathrm{M}_\odot]$
Initial separations r_{12}	1000 AU
Relative velocity $v_{\rm rel}$	2 km/ s
Eccentricity e	$f(e) \sim e$
$e_{\rm min}, e_{\rm max}$	0, 0.99
Impact parameter b	$f(b) \sim b^2$
$b_{\rm min}, b_{\rm max}$	$2,\ 220 - 230$ AU
Semimajor axis (a_1, a_2)	~ 3.2 AU, ~ 1.6 AU
Code termination time	$t = 4$ Myr (for runs with $V_{\rm bg}$)
Stellar radii (binary 1, binary2)	$[16\ \mathrm{R}_\odot + 6\ \mathrm{R}_\odot]$, $[11\ \mathrm{R}_\odot + 8\ \mathrm{R}_\odot]$ (only for runs without $V_{\rm bg}$)
Density ρ	$1\bar{\rho}$, $10^3\bar{\rho}$
Total background mass $M_{\rm bg}$	$3 \times 10^3\ \mathrm{M}_\odot$, $10^4\ \mathrm{M}_\odot$

Table 1: The initial conditions of our scattering experiments. The masses and radii of the four stars are chosen assuming them as proxies for ι Ori, AE Aur and μ Col. These runaways are believed to have formed during encounters in the Trapezium cluster. Note that in simulations with a background potential we do not take into account the stellar radii and physical collisions, while in those without the background potential we explore the frequency of the physical collisions between stars, given the stellar radii shown above. $\bar{\rho} = 1.66 \times 10^{-10}$ g cm^{-3} $= 2450\ \mathrm{M}_\odot\ \mathrm{pc}^{-3}$.

star-forming regions (Blaauw, 1961; R. C. Stone, 1979). However, velocity measurements and reconstruction of trajectories for a number of these stars have shown that they likely originated in stellar clusters, and hence they were ejected from them at high speeds (i.e. Hoogerwerf et al. 2000).

In this study, we examine the hypothesis of the dynamical ejection scenario, namely, dynamical formation as a result of scattering in a close encounter between stars in a star cluster (e.g. Poveda et al., 1967; Gies & Bolton, 1986). We choose the parameters of the initial binaries to match the observed properties of ι Ori binary and two runaway stars (AE Aur and μ Col), which are believed to have formed during a binary-binary encounter in the Trapezium cluster. Furthermore, we focus on one particular case where the end products of the encounters are a binary consisting of ι Ori A and ι Ori B and two single stars. We assume one of the binaries (binary 1) consists of two stars with masses $m_{11} = 39\ \mathrm{M}_\odot$ (a proxy for ι Ori A) and $m_{12} = 16\ \mathrm{M}_\odot$ (AE Aur), while the other binary (binary 2) is composed of two stars with masses $m_{21} = 19\ \mathrm{M}_\odot$ (ι Ori B) and $m_{22} = 16\ \mathrm{M}_\odot$ (μ Col). These specific couplings and the masses of the initial binary components are motivated by Gualandris et al. (2004) (see their Table 2).

We consider two different choices for the stellar radii, zero-size (point particles, or stellar radii $R_\star = 0$) and finite-sized spherical particles. In the simulations with a background potential we only consider point particles, whereas we consider both cases in the simulations

without a background potential. The stellar radii are taken from Table 2 in Gualandris et al. (2004), i.e., $R_{\star,11} = 16$ R$_\odot$, $R_{\star,12} = 6$ R$_\odot$, $R_{\star,21} = 11$ R$_\odot$ and $R_{\star,22} = 8$ R$_\odot$. Gualandris et al. (2004) found that their results have a weak dependence on the choice of the stellar radii. In the finite-size case, we take into account physical collisions between stars. We assume that a physical collision happens when the radii of the stars overlap. Collisions are done in the "sticky star" approximation. Here, stars are treated as rigid spheres with radii equal to their stellar radii. When the radii of two stars overlap, they are merged together with no mass loss and assuming conservation of linear momentum. After the collision, we assume that the radius of the product is equal to the sum of the colliding stars' radii.

We conservatively take the initial separation of the two binaries to be $r_{12} = 1000$ AU and a relative initial velocity of $v_{\rm rel} = 2$ km/ s, which is the mean observed dispersion velocity in the Trapezium cluster (Herbig & Terndrup, 1986). The impact parameter b is randomly drawn from a distribution $f(b) \sim b^2$ within the range $[b_{\rm min}, b_{\rm max}] = [0, 220 - 230$ AU$]$, while the eccentricities of the binaries are randomly generated from a thermal distribution $f(e) \sim 2e$. The mutual inclinations between the binary orbital planes, as well as their initial phases, are randomly chosen.

We adopt the same background potential model in §2.6.1. We consider direct encounters of two binaries with different masses but the same binding energy in the presence/absence of a background potential. We define the ratio of binding energies of the two initial binaries as $\alpha = E_{\rm b,1}/E_{\rm b,2}$, where $E_{\rm b,1}(E_{\rm b,2})$ is the binding energy of the more (less) massive initial binary. This choice is motivated by Gualandris et al. (2004), who find that the distributions of velocities and semimajor axes do not change significantly for $\alpha \leq 3$. They also find that the formation of the ι Ori binary considered in this study is favored by such low ratios of α. We consider two values for the density of the background potential ($\rho = 1$ and 10^3 in units of $\bar{\rho}$, Pfalzner 2009) and two values for the total mass of the background matter ($M_{\rm bg} = 3 \times 10^3$ M$_\odot$ and 10^4 M$_\odot$). These chosen values are typical for young star cluster cores (Portegies Zwart et al. 2010 and references therein), such as e.g. 30 Doradus cluster.

Given the masses above, the total energy of the whole system, $E_{\rm present}$, is estimated from the observed velocities and the relative positions of the four stars (ι Ori binary, AE Aur and ι Ori B, Turon et al. 1992), $E_{\rm present} \simeq -(2-4) \times 10^{48}$ erg. In calculating the total energy, we include the contribution of the background potential $V_{\rm bg}$ (Equation 29) with total mass $M_{\rm bg}$ and with its origin coinciding with the center of mass frame of the four stars. Finally, by energy conservation, the total initial energies in the CM of the four stars can be written as follows,

$$E_{\rm present} = \frac{1}{2}\mu v_{\rm rel}^2 - \frac{G m_1 m_2}{r_{12}} + E_{\rm b,1} + E_{\rm b,2} + \frac{2}{3}\pi G\rho[m_1(r_1^2 - 3r_{\rm bg}^2) + m_2(r_2^2 - 3r_{\rm bg}^2)]\,, \quad (32)$$

where μ is the reduced mass of the two binaries, $\mu = 55 \times 35/(55+35)$ M$_\odot \simeq 21.3$ M$_\odot$ and $E_{\rm b,1}(E_{\rm b,2})$ is the binding energy of binary 1 (binary 2). Also, r_1 (r_2) in the last term (the background gravitational potential for the two binaries) is the distance from the origin to binary 1 (binary 2). In the CM frame, $r_1 = m_2/(m_1+m_2)r_{12}$ and $r_1 + r_2 = r_{12}$. Given

the estimated binding energy reservoir, the semimajor axes of the two binaries are chosen such that their binding energies are equal, $a_1 \simeq 3.2$ AU and $a_2 \simeq 1.6$ AU. Most of the contribution to the total energy comes from the total binding energy $E_{\rm b}$ $(= E_{\rm b,1} + E_{\rm b,2})$ so that the semimajor axes for the two binaries have weak dependences on ρ and r_{12}. Each simulation is run for 4 Myr. This time is chosen to be comparable to and within the estimated uncertainties of the age of a young cluster (e.g. the age of the Trapezium cluster is $\sim 2.5 - 3$ Myr). The initial conditions of our scattering experiments are given in Table 1.

Model	ρ	$M_{\rm bg}$	$r_{\rm bg}$	$\alpha = E_{\rm b,1}/E_{\rm b,2}$	$[b_{\rm min},\ b_{\rm max}]$	$R_\star$	collisions?	r_{12}	$v_{\rm rel}$	$v_{\rm esc}(r = r_{\rm bg})$
Model 0	0	0	0	1	[0, 230 AU]	0	no	$3000 - 6000$ AU	2 km s^{-1}	-
Model 0-1	0	0	0	1	[0, 230 AU]	non-zero	yes	$3000 - 6000$ AU	2 km s^{-1}	-
Model 1	1	3×10^3 M$_\odot$	0.66 pc	1	[0, 230 AU]	0	no	1000 AU	2 km s^{-1}	6.2 km s^{-1}
Model 1-1	1	3×10^3 M$_\odot$	0.66 pc	2	[0, 230 AU]	0	no	1000 AU	2 km s^{-1}	6.2 km s^{-1}
Model 1-2	1	3×10^3 M$_\odot$	0.66 pc	0.5	[0, 230 AU]	0	no	1000 AU	2 km s^{-1}	6.2 km s^{-1}
Model 2	1	10^4 M$_\odot$	1.0 pc	1	[0, 230 AU]	0	no	1000 AU	2 km s^{-1}	9.3 km s^{-1}
Model 3	1000	3×10^3 M$_\odot$	0.066 pc	1	[0, 230 AU]	0	no	1000 AU	2 km s^{-1}	20 km s^{-1}

Table 2: Model parameters used in our study. (From the first column to the last) the model name, the density ρ (in units of $\bar{\rho}$), the total background mass $M_{\rm bg}$, the width of the potential ($r_{\rm bg}$), the ratio of the binding energies of the two initial binaries (α), the range of the impact parameter b drawn from the distribution of $f(b) \sim b^2$, the stellar radius ($R_\star$, Gualandris et al. 2004), the inclusion of the physical collisions, the initial separation r_{12}, the initial relative velocity $v_{\rm rel}$ and the escape velocity $v_{\rm esc}$ at $r = r_{\rm bg}$.

In the simulations without the background potential, we use the same initial conditions described above except that r_{12} is set to be larger by a factor of a few. We refer to this set of simulations as Model 0. We consider three sets of simulations with a background potential given different choices of $M_{\rm bg}$ and ρ (or, simply, the escape velocity $v_{\rm esc}$). We refer to these as Model 1, Model 2 and Model 3. The specific parameters for each of these models are summarized in Table 2. In addition to those four main models, we run simulations with different stellar sizes (Model 0-1) and binding energy ratio (Model 1-1 and 1-2) for investigative purposes. However, in order to avoid adding complexity in interpreting our results, we concentrate on the four main models (without "-1" or "-2") when it comes to the orbital parameters of the interaction.

Based on the results from our scattering experiments, focusing on the formation of the most massive binaries (which will be denoted by $[S_{11} + S_{21}]$ binary) and two ejected single stars (S_{12} and S_{22}), we describe the statistical properties of the simulated final binaries and single stars.

Figure 4 shows the distributions of the semimajor axes a for the $[S_{11} + S_{21}]$ binaries. In the plots for Model 1 to Model 3, we show three different distributions with different line types: those for the binaries which have escaped from the potential before $t = 4$ Myr (dotted blue line), those which have not escaped and remain bound to the potential (thick red solid line), and the overall distributions (thin black solid lines) as the sum of the two distributions. The vertical gold lines indicate an observed value of $a \simeq 0.7$ AU or a period of 29 days (Gualandris et al., 2004).

We can see how, in going from Model 1 to Model 4, the population of final binaries splits into two separate populations, i.e., E-binaries and NE-binaries. The NE-binary population gradually emerges. This is because, as the background potential becomes deeper, the escape velocity increases and hence it gets harder for the stars to escape. The two populations become comparable in size in Model 3, hence contributing equally to the overall distribution. Given the different values in the peaks of the two populations, the overall distribution (thin black solid line) becomes broader. The median values of a for Model 0 are $a_{\rm median} = 1.27$ AU. Those for (NE,2E) case (from Model 1 to Model 3) are $a_{\rm median} = 1.17$ AU, 1.17 AU and 0.94 AU and for (E,2E) case and $a_{\rm median} = 1.73$ AU, 1.62 AU and 1.41 AU. We present the all of the median values for the distributions for the binaries and the single stars in Table 3.

We also notice that the distributions of NE-binaries (red thick solid line) are located at larger a than those of E-binaries. This can be understood in terms of conservation of energy along with the escape velocity. In general, when two single stars are ejected, the recoiled binary carries some kinetic energy. As the recoil velocity of the binary increases, given a fixed total energy, a larger reservoir of negative energy is left for the binary itself, implying a tighter binary. Correspondingly, the fact that binaries could not escape from the background potential means that the instantaneous velocities (or the kinetic energies) of the binaries at the last ejection event were not sufficiently high. Therefore, with smaller energy reservoirs given to the binaries, their semimajor axes are distributed at larger a.

In Figure 5, we present the distributions of the relative angle ξ as a function of the speeds v for Model 0 to Model 3. Here, we define the relative angle ξ as the relative angle between the velocity vectors of the two stars. In the first two circular plots, the distributions for ξ are projected on to the (v, ξ) plane with the median values (magenta pentagon dots). The ξ values for the single stars (the binaries) are marked in the left-half (right-half) of the circle.

Model (Figure)	0	1	2	3
a [AU] (4)	1.27	1.73 / 1.17	1.62 / 1.17	1.41 / 0.935
$v_{\rm binary}$ [km s^{-1}]	22.5	2.24 / 22.8	2.52 / 22.7	5.81 / 25.5
$v_{\rm fast}$ [km s^{-1}]	84.1	35.2 / 92.7	50.4 / 92.2	64.7 / 113
$v_{\rm slow}$ [km s^{-1}]	40.9	28.6 / 46.5	31.1 / 41.8	33.9 / 58.3
$v_{\rm fast}/v_{\rm slow}$	1.86	1.20 / 1.78	1.39 / 1.96	1.71 / 1.89
$v_{\rm fast}/v_{\rm binary}$	3.72	18.4 / 3.92	21.7 / 4.14	10.8 / 4.52
$v_{\rm slow}/v_{\rm binary}$	1.83	14.9 / 1.97	14.4 / 1.98	5.41 / 2.30
$\xi_{\rm fast,slow}$ [$^\circ$]	111	109	105	89.1
$\xi_{\rm fast,binary}$ [$^\circ$]	156	154	156	156
$\xi_{\rm slow,binary}$ [$^\circ$]	105	108	112	120
$\log[t_{\rm bg,\ fast}/{\rm yr}]$	-	3.88	4.04	3.02
$\log[t_{\rm bg,\ slow}/{\rm yr}]$	-	4.16	4.37	3.19
$\log[t_{\rm bg,\ binary}/{\rm yr}]$	-	4.45	4.59	3.39
$\log[t_{\rm bg,\ slow-fast}/{\rm yr}]$	-	3.77	3.94	2.59
$\log[t_{\rm bg,\ binary-fast}/{\rm yr}]$	-	4.29	4.44	3.11

Table 3: The median values of the distributions for the binaries and the single stars. For the models with the background potential (Model 1 to Model 3), we present the median values for (NE, 2E) before "/" and (E, 2E) after "/", separately. Each row (from top to bottom) represents as follows: [Row 1] model name, [Row 2] the semimajor axis a, [Row 3-5] the speeds of the binaries, $S_{\rm fast}$ and $S_{\rm slow}$, [Row 6-8] the speed ratios of $S_{\rm fast}$ to $S_{\rm slow}$, $S_{\rm fast}$ to the binary and $S_{\rm slow}$ to the binary, [Row 9-11] the relative angle ξ between two single stars, between $S_{\rm fast}$ and the binary and between $S_{\rm slow}$ and $S_{\rm fast}$, [Row 12-16] the escape time $t_{\rm bg}$ in $\log_{10}$-scale of $S_{\rm fast}$, $S_{\rm slow}$ and the binary and the escape time of $S_{\rm slow}$ and the binary with respect to $S_{\rm fast}$.

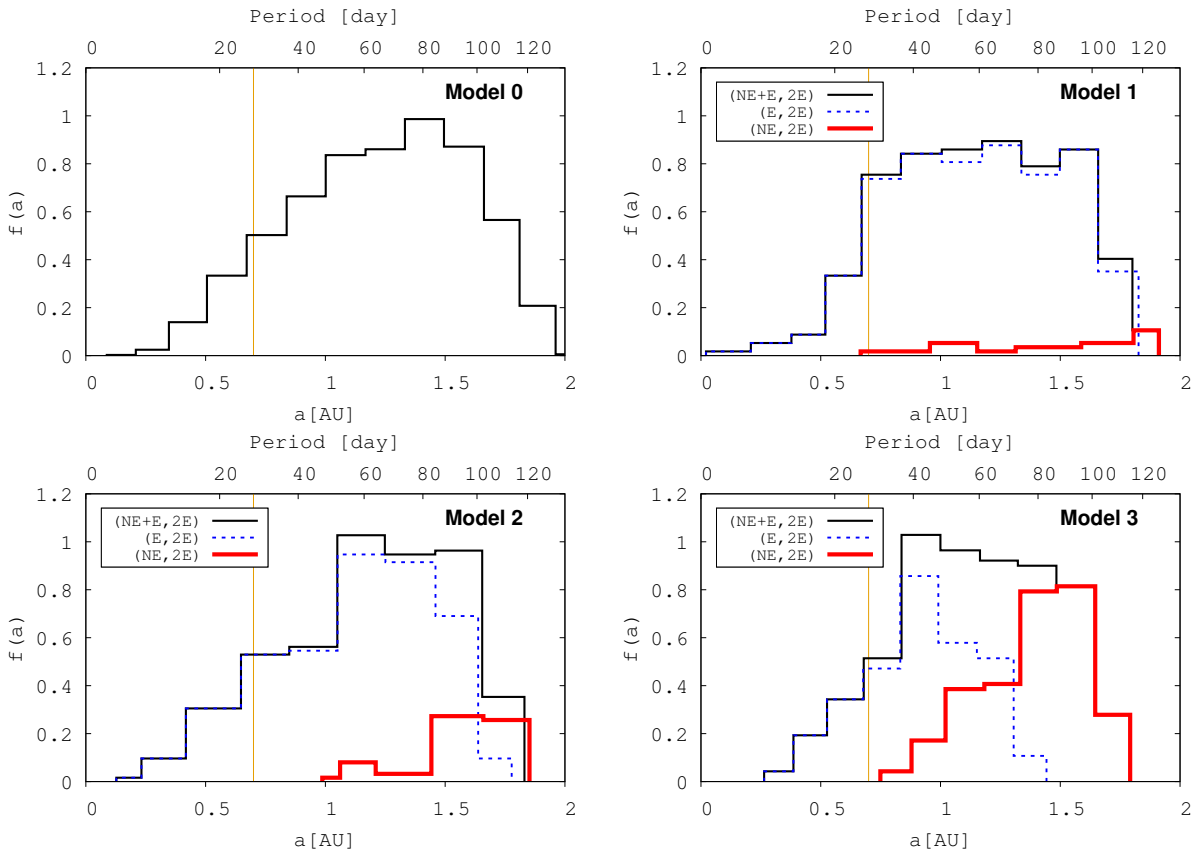

Figure 4: The distributions of the semimajor axes a for the $[S_{11} + S_{21}]$ binaries. The vertical gold lines indicate an observed value of $a \simeq 0.7$ AU or a period of 29 days. In the plots for Model 1 to Model 3, we show three different distributions with different line types: those for E-binaries (dotted blue line) and NE-binaries (thick red solid line) and the overall distributions (thin black solid line) as the sum of the two distributions.

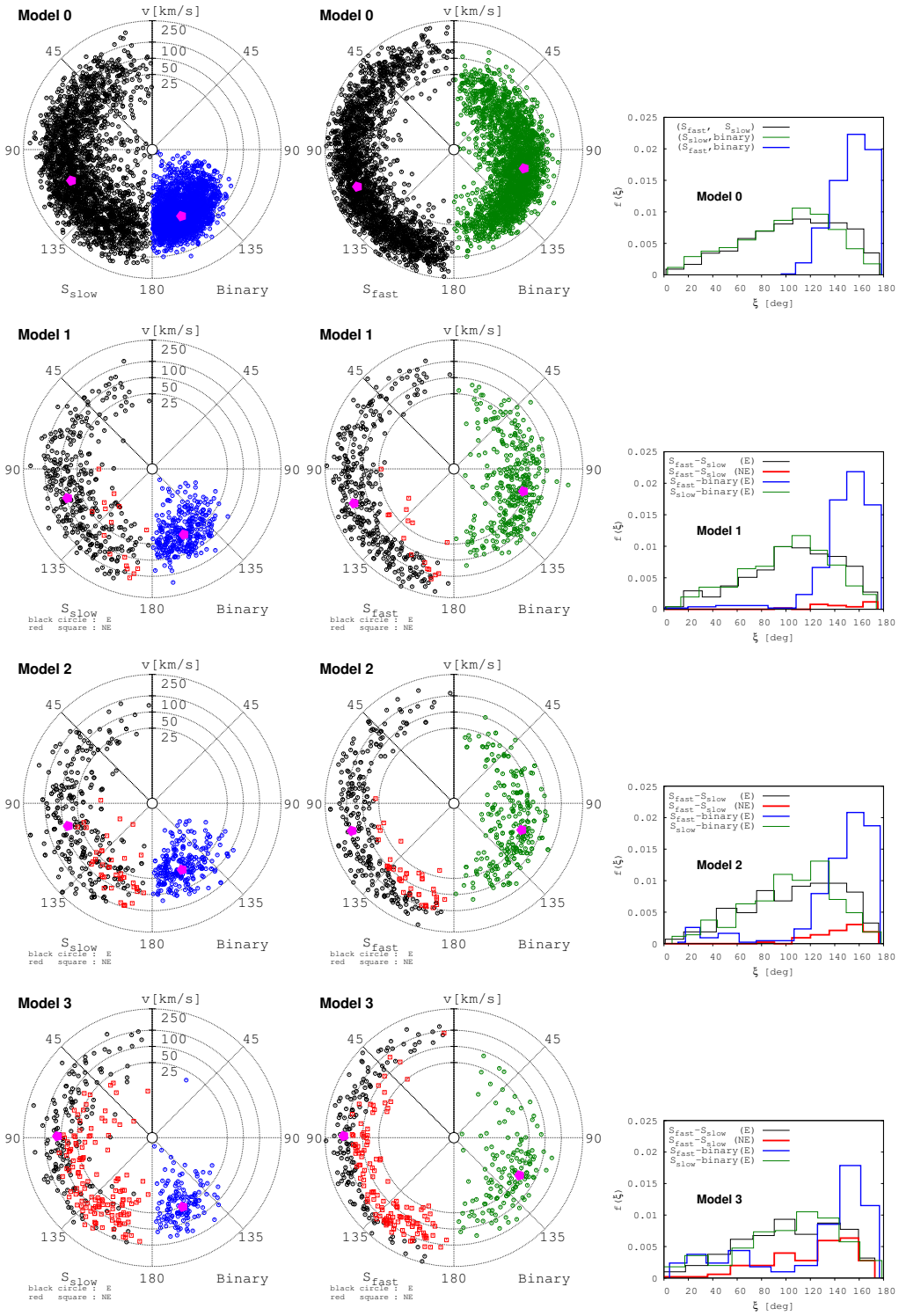

Figure 5: The distributions of the angle ξ between the two stars (single/binary stars) as a function of the speeds v for Model 0-3 (see the definition in the text). In the first two circular plots, the distributions are projected on to the plane (v, ξ) with their median values (magenta pentagon dots). In particular, in the *left* panel, we present ξ of $S_{\rm slow}$ and the binaries with respect to $S_{\rm fast}$ (so v represents the speeds of $S_{\rm slow}$) while in the *middle* panel, we present ξ of $S_{\rm fast}$ and the binaries with respect to $S_{\rm slow}$. And ξ of the single stars (the binaries) are marked in the left-half (right-half) panel. Each dotted circular grid (from the inner circle to the outer circle) indicates velocities of 25, 50, 100 and 250 km s^{-1}. The radial grids (from north to south) show ξ separated by 45° from 0° to 180°. In the *right* panel, we show the distribution functions $f(\xi)$ for ξ. We have used the same colors for the same type of angles in all three panels. In these plots for Model 1-3 (from second low below), we make a distinction between the cases where the binaries have escaped (E, marked with the black circles) and where the binaries have not escaped (NE, marked with the red squares).

In particular, in the *left* panel, we present ξ for $S_{\rm slow}$ and the binaries with respect to $S_{\rm fast}$ (so v represents the speeds for $S_{\rm slow}$ and the binaries) while in the *middle* panel, we present ξ for $S_{\rm fast}$ and the binaries with respect to $S_{\rm slow}$ (so $v = v_{\rm fast}$ and $v_{\rm binary}$). Each dotted circular grid (from the inner circle to the outer circle) indicates velocities of 25, 50, 100 and 250 km s^{-1}. The radial grids (from north to south) show ξ separated by 45° from 0° to 180°. In the *right* panel, we show the distribution function $f(\xi)$ for ξ. We have used the same colors for the same types of angles in all three panels. For example, the blue dots in the *left* panel refer to ξ between $S_{\rm fast}$ and the binaries, which corresponds to the distribution with the blue line in the *right* panel.

In these plots for Model 1-3 (from the second row below), different from Model 0, we make a distinction between cases where the binaries have escaped (E, marked with the black circles) and cases where the binaries have not escaped (NE, marked with the red squares).

There are several characteristic features found from the distributions for ξ in both Figure 5.

1. The relative angles between $S_{\rm fast}$ and the binaries (blue solid lines and blue dots) are densely distributed at $\xi \gtrsim 120° - 135°$ with peaks near $\xi \simeq 160°$.

2. For cases with NE-binaries in Figure 5, the relative angles between the two single stars are more concentrated at higher ξ (red square dots). In addition, as the escape velocity $v_{\rm esc}$ increases (from Model 1 to Model 3), the red dots spread out over a wider range in ξ.

3. The relative angles ξ for $S_{\rm fast}$ and the binaries with respect to $S_{\rm slow}$ are broadly distributed (black/green dots and lines) compared to the distribution for ξ between $S_{\rm fast}$ and the binaries (blue lines and dots), but more biased toward higher values of ξ with peaks at $\xi \simeq 110° - 120°$.

Those features above can be understood in terms of the relation between the recoil velocities of the binaries and the relative angle between the two single stars. If the two single stars are ejected with a wider angle (large ξ), it is more likely that the binary gets a smaller recoil kick. As an extreme case, when two single stars are ejected at the same speed in opposite directions (or $\xi = 180°$), the final recoil velocity of the binary sums to zero (as a result of the two kicks in opposite directions) so that it remains where the last ejection event occurred.

We find feature 1 in our simulations as a result of the fact that kicks from more rapidly-moving stars tend to contribute more to the final recoil velocity of the binary than slowly moving stars. In order to better understand this feature quantitatively, we use Equation 27 from Ryu et al. (2017b) (with $r_{\rm ej} \simeq 0$), which relates the speeds of the binary to the relative angle between the two single stars in a harmonic potential. The equation gives the relative angles $\xi_{\rm fast,slow}$ between $v_{\rm fast}$ and $v_{\rm slow}$ at the outer boundary of the potential as follows,

$$\mathcal{P}_{\rm binary} = \sqrt{\mathcal{P}_{\rm fast}^2 + \mathcal{P}_{\rm slow}^2 + 2\mathcal{P}_{\rm fast}\mathcal{P}_{\rm slow}\cos\xi_{\rm fast,slow}}\ , \tag{33}$$

where $\mathcal{P}_{\rm i} = m_{\rm i}\sqrt{(1/2)v_{\rm esc}^2 + v_i^2}$ (i=binary, $S_{\rm fast}$ or $S_{\rm slow}$) [6]. However, in deriving this equation, the binary and the ejected single star are only distinguished by their masses. Therefore, the

[6]Here, v_i is the speed at $r = r_{\rm bg}$.

relative angles between $S_{\rm fast}$ and the binary $\xi_{\rm fast,binary}$ can be estimated by exchanging $\mathcal{P}_{\rm binary}$ with $\mathcal{P}_{\rm slow}$. Expressing ξ in terms of the momenta,

$$\cos\xi_{\rm fast,binary} = \frac{\mathcal{P}_{\rm slow}^2 - \left[\mathcal{P}_{\rm fast}^2 + \mathcal{P}_{\rm binary}^2\right]}{2\mathcal{P}_{\rm fast}\,\mathcal{P}_{\rm binary}}\ . \tag{34}$$

We can first see that $\xi \geq 90^\circ$ since the numerator is negative. Note that the case for $v_{\rm esc} = 0$ corresponds to Model 0. In particular, for binaries and single stars which have escaped from the background potential, and given the velocity distributions for the binaries and the single stars and their ratio, we can roughly estimate that $\xi_{\rm fast,binary} \simeq 140^\circ$ and $\geq 90^\circ$.

In order to interpret feature 2, we introduce the maximum angle $\xi_{\rm max}$ between two single stars required to give a sufficiently high recoil kick to a binary that the binary can escape to infinity. In other words, if two single stars are ejected with $\xi > \xi_{\rm max}$ at their given speeds, the final binary will not gain enough kinetic energy to completely escape. Therefore, we expect that $\xi_{\rm fast,slow}$ for the NE-binary cases should be distributed at $\xi > \xi_{\rm max}$. We note that $\xi_{\rm max}$ is a function of the speeds of the two single stars ($v_{\rm fast}, v_{\rm slow}$) and the escape velocity $v_{\rm esc}$, namely, $\xi_{\rm max} = \xi_{\rm max}(v_{\rm fast}, v_{\rm slow}, v_{\rm esc})$.

Using $\xi_{\rm max}$, we can understand feature 2 and try to provide useful insight from an observational perspective. Using Equation 33 (or Equation 36 in Ryu et al. 2017b with $r_{\rm ej} \simeq 0$), we impose the condition that at $r = r_{\rm bg}$ and $\xi = \xi_{\rm max}$, $v_{\rm binary} = v_{\rm esc}$. More explicitly,

$$\sqrt{\mathcal{P}_{\rm fast}^2 + \mathcal{P}_{\rm slow}^2 + 2\mathcal{P}_{\rm fast}\mathcal{P}_{\rm slow}\cos\xi_{\rm max}} = \sqrt{\frac{3}{2}} m_{\rm binary} v_{\rm esc}\ . \tag{35}$$

Again, the term on the left-hand side represents the final momentum of the binary at the last ejection event and the term on the right-hand side the minimum momentum of the binary necessary to travel from the origin to the outer boundary of the potential, and subsequently to spatial infinity.

In the *upper* panel of Figure 6, we estimate $\xi_{\rm max}$ (using Equation 35) as a function of $v_{\rm slow}$ in units of $v_{\rm esc}$ for various ratios of $v_{\rm fast}/v_{\rm slow}$. As shown in the figure, $\xi_{\rm max}$ increases rapidly at $v_{\rm slow}/v_{\rm esc} \lesssim 5$. It reaches $120^\circ - 140^\circ$ at typical velocities for $S_{\rm slow}$ found in Model 1 and 2 (median velocities of $v_{\rm slow}/v_{\rm esc} \sim 3$) for all speed ratios, which explains well the feature 2. However, considering smaller values for $v_{\rm slow}/v_{\rm esc} \lesssim 2$ in Model 3, it is hard to read $\xi_{\rm max}$ from the *upper* panel. Hence we additionally provide a plot (*bottom* panel), which shows $\xi_{\rm max}$ as a function of $v_{\rm esc}$ assuming $v_{\rm slow} \sim 45$ km s^{-1}. We can see that for $v_{\rm esc} = 20$ km s^{-1} in Model 3 (indicated with the downward arrow), $\xi_{\rm max}$ decreases down to $40^\circ - 120^\circ$, which accounts for the wide range of dots in Figure 5, or feature 2.

From an observational perspective, we expect that the analytic relations above (as derived in the analytic paper) and the statistical properties of the relative angles found from our simulations can help restrict the region in parameter space where we need to look in order to find an unknown related object given observations of some other runaway stars. Note that, in the derivation of the above relation, the three bodies are only identified by their masses, meaning that the relation can be applied to any kind of isolated system that evolves to produce 3-body outcomes (two ejected systems and one left-over system), not necessarily only to the 2+1+1 outcome.

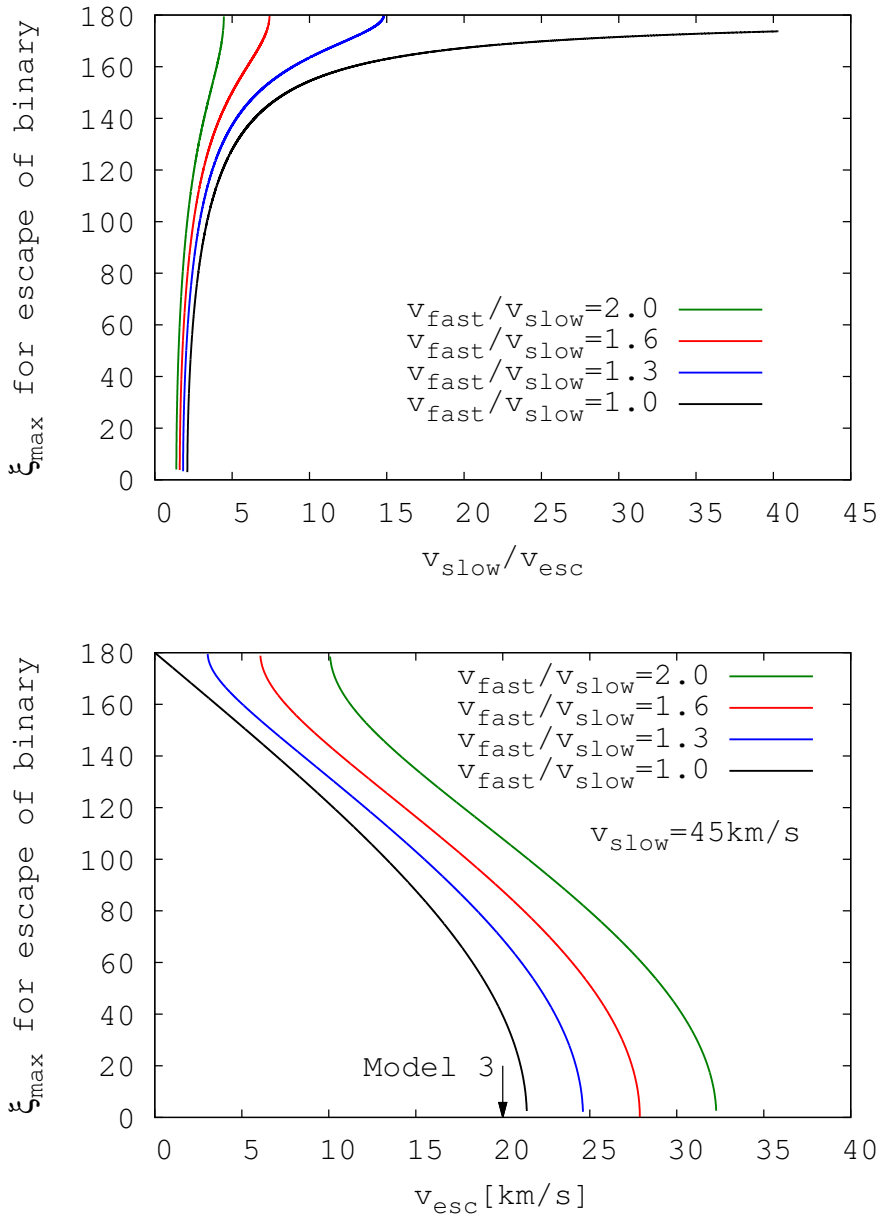

Figure 6: $\xi_{\rm max}$ estimated using Equation 35. *Upper* panel: $\xi_{\rm max}$ is expressed as a function of $v_{\rm slow}$ in units of $v_{\rm esc}$ for various ratios of $v_{\rm fast}/v_{\rm slow}$. $\xi_{\rm max}$ increases rapidly up to $120^\circ - 140^\circ$ at typical velocities for $S_{\rm slow}$ for Model 1 and 2. In the *lower* panel, $\xi_{\rm max}$ is described in terms of $v_{\rm esc}$ assuming $v_{\rm slow} \sim 40$ km s^{-1}. $\xi_{\rm max}$ decreases down to $50^\circ - 90^\circ$ for $v_{\rm esc} = 20$ km s^{-1} in Model 3 indicated by the downward arrow (feature 2).

3 Formation, disruption and energy output of Population III X-ray binaries

T. Ryu, T. Tanaka, R. Perna *The Monthly Notices of the Royal Astronomical Society*, Vol. 456, 223 (2016)

Abstract

The first astrophysical objects shaped the cosmic environment by reionizing and heating the intergalactic medium (IGM). Particularly, X-rays are very efficient at heating the IGM before reionization is complete. High-mass X-ray binaries (HMXBs) in early stellar populations are prime candidates for driving the thermal evolution of the IGM at redshifts $z \gtrsim 20$; however, their formation efficiency is not well understood. Using N-body simulations, we estimate the HMXB formation rate via mutual gravitational interactions of nascent, small groups of the Population III (Pop III) stars. We run two sets of calculations: (i) stars formed in small groups of five in nearly Keplerian initial orbits and (ii) collision of two such groups (an expected outcome of mergers of host protogalaxies). We find that HMXBs form at a rate of one per $\gtrsim 10^4$ $M_\odot$ in newly born stars, and that they emit with a power of $\sim 10^{41}$ erg s^{-1} in the $2-10$ keV band per star formation rate (SFR). This value is a factor $\sim 10^2$ larger than what is observed in star forming galaxies at lower redshifts; the X-ray production from early HMXBs would have been even more copious, if they also formed *in situ* or via migration in protostellar disks. Combining our results with earlier studies suggests that early HMXBs were highly effective at heating the IGM and leaving a strong 21 cm signature. We discuss broader implications of our results, such as the rate of long gamma-ray bursts from Population III stars and the direct collapse channel for massive black hole formation.

3.1 Introduction

A major outstanding goal in cosmology is to piece together the history of the Universe between Cosmic Dawn, the emergence of the first stars and galaxies, and the end of reionization, when the radiation from these objects had ionized the intergalactic medium (IGM). Advances in numerical techniques, combined with exquisite measurements of the "initial" conditions (at a redshift $z \approx 1000$; Hinshaw et al. 2013, Planck Collaboration et al. 2015), have led to remarkable simulations (e.g. Abel et al., 2002; Turk et al., 2009; Stacy et al., 2010; Greif et al., 2011; Bromm & Yoshida, 2011) of the conditions leading up to the former milestone, occurring at $z \gtrsim 30$, when the Universe was ≈ 100 Myr old. However, reconstructing the subsequent several hundred Myr of cosmic history has proved far more challenging, due to the difficulties in reliably modeling the numerous forms of feedback from the first astrophysical objects (e.g. Springel, Di Matteo, & Hernquist, 2005; Stinson et al., 2006; Sijacki et al., 2007).

In particular, X-rays from the first galaxies can act as a powerful source of feedback (Venkatesan et al., 2001; Machacek et al., 2003) that exerts influence over a wide range of distance scales. Because hard X-rays (energies $\gtrsim 1$ keV) have mean free paths comparable to the Hubble horizon, they can isotropically heat and partially reionize the early IGM (Oh,

2001; Venkatesan et al., 2003; Ricotti & Ostriker, 2004a; Pritchard & Furlanetto, 2007). In fact, they are expected to be the dominant agent in heating the IGM. Such heating may suppress star formation (Ripamonti et al., 2008) and massive BH growth (Tanaka et al., 2012a) inside low-mass dark matter haloes by raising the Jeans and filtering masses of the IGM (Gnedin, 2000; Naoz & Barkana, 2007). On galactic and circum-galactic scales, soft X-rays ($\sim 0.1 - 1$ keV) can affect the formation of stars and possibly massive black holes (BHs) by promoting the formation of molecular hydrogen via electron-catalyzed reactions (Haiman et al., 1996; Kuhlen & Madau, 2005; Latif et al., 2015; Inayoshi & Tanaka, 2015).

In addition to their suspected roles in early galaxy evolution, X-rays are important also because they can leave an observable signature that can be exploited to probe the cosmological epoch in question (Pritchard & Loeb, 2008). Their thermal impact on the early Universe should be measurable through the redshifted 21 cm transition line of neutral hydrogen, which is observed in emission or absorption depending on the relative temperature of the IGM with respect to that of the cosmic microwave background (CMB). Several studies have investigated how forthcoming observations of the sky-average amplitude and power spectrum of the relic 21 cm line (Bowman et al., 2008; Burns et al., 2012; van Haarlem et al., 2013; Voytek et al., 2014) could be used to constrain the astrophysical agent (or agents) responsible for heating the early IGM.

There are sound reasons to expect that the first galaxies produced X-rays in abundance, and rapidly heated the IGM. There are two dominant X-ray sources in present-day galaxies—both powered by gas accretion onto BHs, and both plausibly prominent shortly after Cosmic Dawn: gas feeding massive BHs shining as active galactic nuclei (AGN), and X-ray binaries, powered by a stellar-mass BH gradually cannibalizing a companion star. Estimates of the mass accumulated by nuclear BHs prior to $z \sim 6$ (Shankar et al., 2009; Salvaterra et al., 2012), the existence of very massive BHs at $z \gtrsim 6$ (Fan et al., 2001; McGreer et al., 2006; Willott et al., 2007, 2009; Mortlock et al., 2011; Venemans et al., 2013; Bañados et al., 2014), as well as the observed (Shen et al., 2007) and theoretically expected (Shankar et al., 2009; Tanaka, 2014) increase in their duty cycles toward higher redshifts, all hint that X-ray AGN may have been much more common during this epoch. Likewise, high-mass X-ray binaries (HMXBs) dominate the X-ray emission of star-forming galaxies; the low metallicity and rapid baryonic mass accretion of the earliest galaxies both lend credence to the notion that they were rife with HMXBs. Theoretical models suggest that either type of X-ray source could heat the IGM to above the CMB temperature as early as $z \sim 30$, and that this transition should be measurable by the planned 21 cm experiments.

At present, there are too many theoretical uncertainties to determine from the future data which type of X-ray source—AGN or HMXBs—was responsible for driving the thermal evolution of the $z \lesssim 30$ IGM. Modeling the early AGN X-ray emission is particularly difficult, because the conditions for triggering AGN activity are not fully understood even at low redshifts (Hopkins & Quataert, 2010; Treister et al., 2012); because of the uncertainty in the fraction of X-ray photons that is released into the IGM as opposed to being trapped inside the accretion flow or reprocessed into the infrared (Madau et al., 2014; Pacucci, Ferrara, et al., 2015); and because the epoch, initial masses, and birthplaces of the massive BH "seeds" are not yet constrained by observations (Volonteri, 2010; Haiman, 2013; Tanaka & Li, 2014). Similarly, studies usually estimate the X-ray contribution from early HMXBs by simply inferring empirical relations between X-ray luminosity and star formation rate

(SFR) in local galaxies (and modeling the SFR using a semi-analytic cosmological model), or combining such relations with one or more free parameters (Mirabel et al., 2011; Tanaka et al., 2016).

The goal of this study is to alleviate the uncertainties in the formation rate and X-ray output of HMXBs in the early Universe, by using N-body simulations of nascent groups of the first (or Pop III) stars. We choose the properties of the star groups in our simulations to reflect those found in hydrodynamical simulations of Pop III star formation at $z \sim 20$ (Greif et al., 2012; Stacy & Bromm, 2013). We follow the formation and dynamical evolution of compact binaries over thousands to millions of years, including the effects of the background gravitational potential and dynamical friction. This allows us to compute the fraction of Pop III stars that form stable, compact binaries, and eventually undergo an X-ray bright phase. The end result is an estimate of the formation rate of HMXBs in the first protogalaxies, as well as the amount of X-rays they generate per unit star formation. To our knowledge, this is the first published estimate of this type.

We derive a HMXB energy output (normalized to the star formation rate) that is a factor $\sim 10-150$ higher than in present-day star-forming galaxies, if the HMXB duty cycle is similar to the one in the local Universe. We find that the X-ray output does not change significantly within the wide variety of simulation setups considered—such as different orientations for collisions between star groups, and ambient gas density—and submit that this is a robust estimate.

The findings of this study can be used as model inputs in estimating the 21 cm global signature and power spectrum, but have wider applications. As stated above, the X-ray output of the first galaxies are also of interest for studying feedback on smaller scales, such as subsequent star formation and massive BH formation.

Our work is also relevant for predicting the rates of long-duration gamma-ray bursts (LGRBs) from Pop III stars. LGRBs are important probes that can shed light on the Universe out to $z > 10$ (Toma et al., 2011). According to the collapsar model (MacFadyen & Woosley, 1999), progenitors of LGRBs require rapid rotation of the He core and removal of the H envelope. Both criteria are satisfied by Pop III HMXBs, and it is plausible that massive Pop III stars in binary systems are dominant LGRB progenitors in the early Universe (Bromm & Loeb, 2006). Our results on HMXB formation rates can therefore be used to predict and interpret observations of high-redshift LGRBs.

The paper is organized as follows. We start in §3.2 by discussing the problem to be solved—beginning with the equations of motion, followed by the description of our N-body code, our choices for the initial conditions, and how the data is interpreted for HMXB formation. We present our results in §3.3. In §3.4, we discuss the implication of our work for the X-ray output of the first galaxies, as well as for other topics such as LGRBs and SMBH formation. We conclude with a summary of our findings in §3.5.

3.2 Stellar Dynamics

Here, we provide an overview of our simulations—namely: the equations of motion that are solved to simulate the dynamical evolution of the star groups; the numerical scheme we use to solve the equations; the different types of initial conditions we adopted, as well as the reasoning behind our choices; and finally, how the results are interpreted for HMXB

formation.

3.2.1 The equations of motion

Our N-body code computes the motion of N objects with (generally different) masses m_i, moving under their mutual gravitational influence, a dissipative dynamical friction force, and a background gravitational potential. We numerically integrate the equations of motion

$$\frac{d^2}{dt^2}\vec{r}_i = \vec{a}_{\mathrm{g},i} + \vec{a}_{\mathrm{df},i} + \vec{a}_{\mathrm{bg},i}. \tag{36}$$

The first term on the right-hand side of equation (36) is the specific force due to Newtonian gravity,

$$\vec{a}_{\mathrm{g},i} = -\sum_{j\neq i} G\ m_j\ \frac{\partial\ S(r_{ij})}{\partial\ r_{ij}}\ \frac{\vec{r}_i - \vec{r}_j}{r_{ij}}, \tag{37}$$

where G is the gravitational constant, $\vec{r}_i$ is the displacement of the i_{th} star from the center of the host dark matter halo, and $r_{ij} \equiv |\vec{r}_i - \vec{r}_j|$.

We adopt the Plummer softening kernel $S(r_{ij})$ (Binney & Tremaine, 1987),

$$S(r_{ij}) = -\frac{1}{\sqrt{r_{ij}^2 + \epsilon^2}}\,, \tag{38}$$

where we take $\epsilon = R_\odot$.

The second term on the right-hand side of equation (36), $\vec{a}_{\mathrm{df},i}$, is the specific drag force due to dynamical friction. For collisionless systems, the standard Chandrasekhar formula for dynamical friction is (Binney & Tremaine, 1987),

$$\vec{a}_i = -4\pi\ \ln\Lambda\ f(X_i)\ \frac{G^2 m_i}{v_i^3}\ \rho(r_i)\ \vec{v}_i, \tag{39}$$

where

$$f(X_i) \equiv \mathrm{erf}(X_i) - \frac{2}{\sqrt{\pi}}\ X_i\ \exp\left(-X_i^2\right), \tag{40}$$

v_i is the speed of the i_{th} star with respect to the background, $X_i \equiv v_i/(\sqrt{2}\sigma_v)$, σ is the velocity dispersion, $\ln\Lambda$ is the Coulomb logarithm and $\rho(\vec{r}_i)$ is the local gas density.

We adopt the modified formula for gaseous medium used in Tanaka & Haiman (2009). This prescription incorporates behaviors found in numerical simulations for subsonic and supersonic regimes (Ostriker, 1999b; Escala et al., 2004). The specific drag force vector always points opposite to the direction of motion, and is given by:

$$a_{\mathrm{df},i}^{(\mathrm{gas})} = -4\pi\ G^2\ m_i\ \rho(\vec{r}_i)\ \frac{1}{v_i^2} \times f^{(\mathrm{gas})}(\mathcal{M}_i), \tag{41}$$

with

$$f^{(\mathrm{gas})}(\mathcal{M}_i) = \begin{cases} 0.5\ \ln\Lambda\Big[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i\ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\Big] \\ \qquad\qquad 0 \leqslant \mathcal{M}_i \leqslant 0.8; \\ 1.5\ \ln\Lambda\Big[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i\ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\Big] \\ \qquad\qquad 0.8 \leqslant \mathcal{M} \leqslant \mathcal{M}_{\mathrm{eq}}; \\ \frac{1}{2}\ln\left(1 - \frac{1}{\mathcal{M}_i^2}\right) + \ln\Lambda \\ \qquad\qquad\qquad\qquad \mathcal{M}_i > \mathcal{M}_{\mathrm{eq}}. \end{cases} \tag{42}$$

Above, $\mathcal{M}_i \equiv v_i/c_\mathrm{s}$ is the Mach number, and c_s is the sound speed. We use $\ln\Lambda = 3.1$ and the corresponding value of $\mathcal{M}_{\mathrm{eq}} \approx 1.5$.

In our simulations, the motion of the stars with respect to the background gas is supersonic. In this regime, the characteristic dynamical friction timescale, for a circular Keplerian orbit of two bodies with for $m_1 \gg m_2$ and $v_2 \gg c_\mathrm{s}$, is

$$\tau_\mathrm{df} \sim \frac{E_\mathrm{orb}}{P_\mathrm{df}} \sim \frac{1}{80}\sqrt{\frac{m_1^3}{m_2^2 G}}\frac{1}{\rho(r_2) r_{12}^{3/2}}\,, \tag{43}$$

where P_df is the frictional dissipation power ($m_2 a_{\mathrm{df},2} v_2$), and E_orb the orbital energy ($G m_1 m_2/2r_{12}$).

The third and last term, $\vec{a}_{\mathrm{bg},i}$, is the specific force due to the background potential, which is dominated by gas. The background potential provides an additional inward force whose functional form depends on the density profile. For simplicity, here we use a constant density and explore different values in our simulations. The force due to the background potential is then

$$\vec{a}_{\mathrm{bg},i} = -4\pi G\rho\vec{r}_i, \tag{44}$$

where here $\vec{r}_i$ is the vector pointing from the center of the halo to the i-th star.

The equation of motion is solved iteratively, with the positions, velocities and accelerations of each star updated at every time step. We describe our computational method below.

3.2.2 Numerical technique

We implement the following two numerical shortcuts, more specific to this study, to keep computation times tractable.

The first shortcut is to use analytic approximations for very close 2-body encounters. This is justified in cases where pairs of stars are sufficiently close to each other and isolated from the other stars in the simulation, so that (i) the gravitational pull from the other stars and the background potential are negligible compared to the mutual gravitational pull of the pair, and (ii) the orbital motion is supersonic and the dynamical friction force can be treated as a linear perturbation to the 2-body Keplerian problem. The code employs analytic approximations for any close stellar pairs that satisfy these conditions, and reverts to the RKF algorithm when the conditions stop being satisfied.

We derived the following approximations for the semi-major axis and the eccentricity:

$$a(t) = \left[a(t_0)^{-3/2} + \beta \ \rho \ \sqrt{\frac{G}{\widetilde{\mu} w(q)}} \ (t - t_0) \right]^{-2/3} , \tag{45}$$

$$e(t) = \sqrt{1 - \frac{a(t)}{a(t_0)}(1 - e(t_0)^2)} \, . \tag{46}$$

Above, $\widetilde{\mu}$ is the reduced mass, $a(t_0)$ and $e(t_0)$ are respectively the semi-major axis and the eccentricity at $t = t_0$, β is a dimensionless constant, and $w(q)$ is a function of the mass ratio that is symmetric about $q = 1$. Note that the variable a with subscript indicates the specific force, while without subscript it represents the semi-major axis. To derive the analytical expression for the time-evolution of the orbital distance in equation (45), we integrated the equation of the motion of a star under the influence of a dynamical friction torque. Equation (46) then follows from the definition of orbital eccentricity in terms of orbital energy and angular momentum. Note that as the radial distance decreases, the eccentricity increases. Here, β and $w(q)$ are free parameters and we use $\beta = 0.035$ and $w(q) = q^{1.4} + (1/q)^{1.4}$. The stellar coordinates and velocities are recovered as functions of a and e using standard expressions for Keplerian orbits (Binney & Tremaine, 1987).

Our second shortcut for keeping the simulation runtimes manageable is to set a minimum value for the time step. We choose a physically motivated value, $10^{-6} \times \tau_{\rm dyn,\,min}$, where $\tau_{\rm dyn,\,min}$ is the smallest value of the dynamical time between any two stars in the simulation (that are not being treated by the analytic shortcut above) at a given time step. This procedure is necessary when there are three or more stars interacting at small separations, in which case the analytic approximations above cannot be used. We note that such situations are rare compared to the places where the analytic shortcut is applicable.

3.2.3 Determining HMXB formation

It is assumed that a HMXB has formed if both of the following criteria are satisfied:

1. *One of two stars forming a binary turns into a compact object (CO).* In order to determine which stars turn into COs, we need to compare the typical lifetime ($\tau_{\rm life}$) of a massive star with the time ($t_{\rm run}$) in the simulation (taken to coincide with the time at which stars are born). If $\tau_{\rm life} > t_{\rm run}$, the star is marked as a CO in the simulation. For main-sequence stars, it is possible to estimate this lifetime from the mass-luminosity relation, that is $E_{\rm star} \sim m$ and $L_{\rm star} \sim m^p$, leading to $\tau_{\rm life} \sim m^{1-p}$ with $2 < p < 3$. However, due to the uncertain value of p for Pop III stars, we rather prefer to use here the nuclear time scale of Pop III stars estimated by Schaerer (2002) and Marigo et al. (2003). They calculate the H-burning nuclear time scale for these stars with a stellar evolution code. We assume that a CO forms if the stellar mass is greater than 8 $\rm M_\odot$.

2. *The two stars are close enough so that the accretion occurs through Roche-lobe overflow (RLOF).* This criterion is simply written as $R_{\rm star} \geq R_{\rm RL}$, where $R_{\rm star}$ is the stellar radius, and $R_{\rm RL}$ is the Roche-lobe radius of the most massive star calculated from the

center of the star to the inner Lagrange point. An approximate analytic formula to the Roche-lobe radius of star 1 for a wide range of the mass ratio is (Eggleton, 1983),

$$\frac{R_{\rm RL,1}}{r} = \frac{0.49q^{2/3}}{0.6q^{2/3} + \ln(1+q^{1/3})}\,, \tag{47}$$

where r is the orbital distance and $q = \frac{m_2}{m_1}$ is the mass ratio. Although the above expression was derived in a study of binaries in circular orbits, Regös et al. (2005) found that the Roche lobe radius does not differ very much for eccentric binaries. We therefore apply equation (47) to all binaries found in our simulation.

If a binary satisfies these two criteria, the binary is marked as a HMXB. Since the two criteria are independently checked at every time step, which criterion is satisfied first is not important.

3.2.4 Setup and Initial Conditions

We design our simulations with the $20 \lesssim z \lesssim 30$ Universe in mind. This is the redshift range where the dark matter haloes reach virial temperatures $\sim 1000 - 2000$ K, the expected condition for Pop III formation, at the highest rates.

The lifetimes of the stars are several Myr, which is comparable to the timescales on which the host haloes undergo mergers with other Pop III forming haloes (Tanaka, 2014).

The initial positions of the stars are generated quasi-randomly via a Monte Carlo realization as follows, motivated by the assumption that they formed inside a shared Keplerian gas disk at the center of the host dark matter halo. Their initial radial positions in the disk plane are chosen randomly from a uniform distribution whose amplitude depends on the characteristic size scale of the star-forming cloud (see below). Their azimuthal distribution is chosen to be nearly uniformly distributed, so that the azimuthal angular position of the n-th star is given by $360°(n/N) \pm 5°$, where N is the total number of stars in the group. This choice is made to minimize the radial gravitational pull between the stars, so that they do not immediately fly apart. Naturally, the azimuthal positions do not remain evenly spaced, but become mixed quickly. Finally, we allow the stars to be displaced out of the plane of the disk. Their positions perpendicular to the plane are chosen randomly, so that their vertical displacement out of the disk plane is no more than 5% above or below their initial radial displacement from their shared center of mass. The initial velocities of the stars are assigned to be circular, parallel to the disk plane, and Keplerian at the instant the simulation begins.

For this study, we investigated two different size scales of star forming regions.

1. *Large scale*: Following the results from Stacy & Bromm (2013), where Pop III protostars are formed inside star-forming regions with a size of a few thousand AU, we have performed simulations where the stars are placed inside a region of size ~ 2000 AU.

   We explore two different values of uniform, constant gas density, 10^6 cm^{-3} (denoted n_6 hereafter) and 10^4 cm^{-3} (n_4).

   For gas of primordial composition, the molecular weight μ can vary between 0.6 and 1.2, depending on the ionization fraction. We adopt $\mu = 1$ for simplicity; this choice

does not qualitatively affect our results, since our values for n are selected arbitrarily with the goal of exploring the qualitative dependence on n.

Each simulation is run for 5 Myr. The masses of the stars follow the initial mass function (IMF) they provide with $\alpha = 0.17$ ($\frac{dN}{dM} = M^{-\alpha}$), $M_{\rm max} = 140$ $M_\odot$ and $M_{\rm min} = 0.1$ $M_\odot$. For these simulations, we consider $N = 5$ stars.

2. *Small scale*: In the simulations by Greif et al. (2012), multiple protostars formed several AU apart from each other. We therefore run a second set of simulations, where stars form within a 10 AU radius. Since Greif et al. (2012) do not provide a slope for the IMF, we use $\alpha = 0.17$ as above. That study found that more than half of the mass accreted during the protostar phase goes to the most massive protostar in the group.

   To mimic this behavior, we first generate a star with $M_{\rm max} = 200$ $M_\odot$, then generate each subsequent star with $M_{\rm max} = 200$ $M_\odot -$ [the sum of the masses of the previously generated stars].

   Just as with the large-scale case, we run simulations with number densities n_6 and n_4. Because this case is the more relevant one for forming HMXBs, we explore several different configurations: cases with a star group of $N = 5$ stars, a star group of $N = 10$ stars, and collisions between two groups of $N = 5$ stars. The last case is motivated by the fact that Pop III-forming minihaloes undergo frequent mergers, which suggests that the nascent star groups themselves undergo close encounters.

3.3 Results

Here, we summarize the findings of our N-body simulations, focusing in particular on the properties of the most compact binaries found for each set of runs.

3.3.1 Large scale

We ran 20 simulations for n_6 and 23 for n_4.

For n_6, in 12 out of 20 runs, the two most massive stars (S_1 and S_2, where S_i is the i-th most massive star) form the most compact binary—we denote such a binary with the notation B_{12}. Similarly, binaries of type B_{13} (i.e. made up of the most massive star S_1 and the third most massive star S_3) form the most compact binary in 6 of the runs.

The left panel of Figure 7 shows a sample set of stellar trajectories for one of the large scale n_6 runs. Throughout the simulation, the compact binary tends to remain near the center of the halo, as less massive stars repeatedly undergo 3-body interactions with the binary. The fact that they do not stray far from the center of the halo is a combined effect of the background potential and the gravitational potential of the massive binary. It is these 3-body encounters that cause the most compact binary in the simulations to end up as type B_{12} or B_{13}. In two of the runs, the most compact binaries after 5 Myr consist of less massive stars.

The average value of the semimajor axis after $t = 5$ Myr in the large scale, n_6 runs is $\langle a_{t=5\ {\rm Myr}} \rangle = 270$ AU, and the minimum value of a across the 20 runs is 60 AU. These values for a are much larger than that necessary for RLOF to take place, ~ 0.07 AU. Their

Large scale	n_6	n_4
number of runs	20	23
$P(\mathrm{B}_{12})$ $[P(\mathrm{B}_{12}+\mathrm{B}_{13})]$	0.60(0.90)	0.56(0.74)
$\langle a_{t=5\mathrm{Myr}}\rangle$[AU]	270	340

Table 4: Summary of the large scale calculations for $n = 10^6$ cm^{-3} (n_6) and $n = 10^4$ cm^{-3} (n_4). $P(\mathrm{B}_{12})$ denotes the fraction of the simulations in which the most compact binary consists of the two most massive stars. We also list $P(\mathrm{B}_{12} + \mathrm{B}_{13})$, the fraction where the most compact binary consists of the most massive star paired with either the second or third most massive star. Also shown is the average semi-major axis of the most compact binary at $t = 5$ Myr.

characteristic dynamical friction time scales are roughly 10^{13} yrs, which is much longer than their lifetimes (see equation 47 when $R_{\mathrm{RL}} = R_{\mathrm{star}}$ and equation 43). We therefore conclude that in systems of $N = 5$ stars and $n \sim 10^6$ cm^{-3}, with the stars initially separated at hundreds of AU, HMXBs are unlikely to form.

For the n_4 case, the fraction of runs where the most compact binary after 5 Myr is type B_{12} or B_{13} are similar to the n_6 simulations: 13/23 for B_{12}, and 17/23 for B_{12} *or* B_{13}.

However, the dynamical evolution is quite different, as can be seen by the right panel of Figure 7. The lower gas densities lead to weaker forces due to dynamical friction and background potential, and stars (especially less massive ones) tend to be scattered farther from their initial position, as far as $\gtrsim 1$ pc. This also results in less frequent three-body interactions in general, and may explain the larger average value of the semi-major axis after 5 Myr, $\langle a_{t=5\mathrm{Myr}}\rangle = 340$ AU. The formation of HMXBs appears even more intractable for the n_4 case.

Sample stellar trajectories for the n_6 and n_4 large scale cases are shown in Figure 7, and a summary of the results is given in Table 4.

3.3.2 Small scale

For the small scale case, we performed 86 runs of 5-body simulations for each of the number density values n_6 and n_4. Due to the smaller initial separations of the stars, we run the simulations for a shorter amount of time.

We run each simulation for a minimum of 1000 yr, but stop the run if a stable binary forms, and if no further significant dynamical changes are observed. If such a binary does not form, we run the simulations to a maximum duration of 5000 yr.

In order to properly compare the results from the n_6 calculations with those from the n_4 calculations, we use the same initial conditions for each set of runs.

In the n_6 calculations, the most common scenario is that S_1 always forms the most compact binary almost immediately, while stellar scatterings are most common during the first few years to about 40 years. Thereafter, a multiple system usually survives and stabilizes, while less massive stars are ejected. A difference between the large-scale and the small-scale scenarios is that, whereas most cases in the large-scale calculations end up with one binary

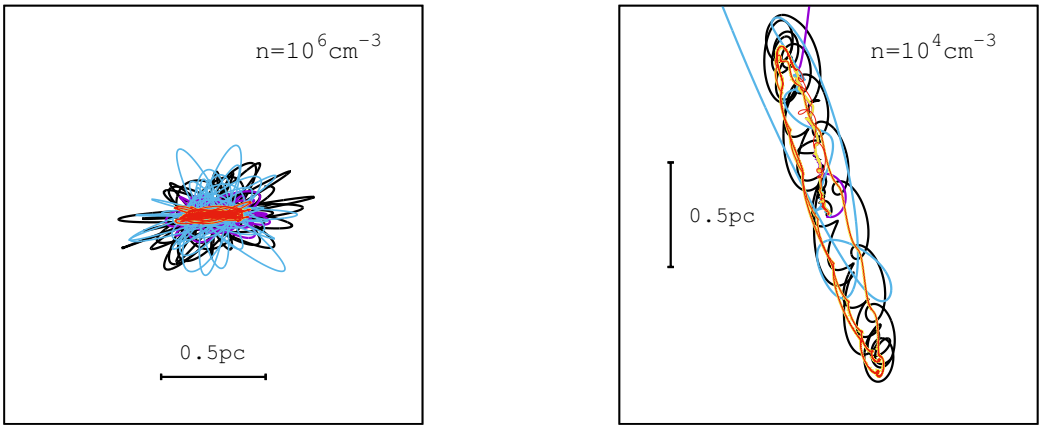

Figure 7: *Left*: Sample trajectories for large scale calculations with the high number density $n = 10^6$ cm^{-3} (case n_6). As shown in the figure, stars tend to stay near the center of the halo and their overall motions are oblate-spheroidal in shape. *Right*: Sample trajectories for the low number density $n = 10^4$ cm^{-3} (n_4). Even though stars including binary systems remain within a certain distance range, they are not as close as the stars in the higher density calculation, leading to less frequent three-body interactions. Furthermore, one can notice that a few stars (blue line and purple line) are kicked off.

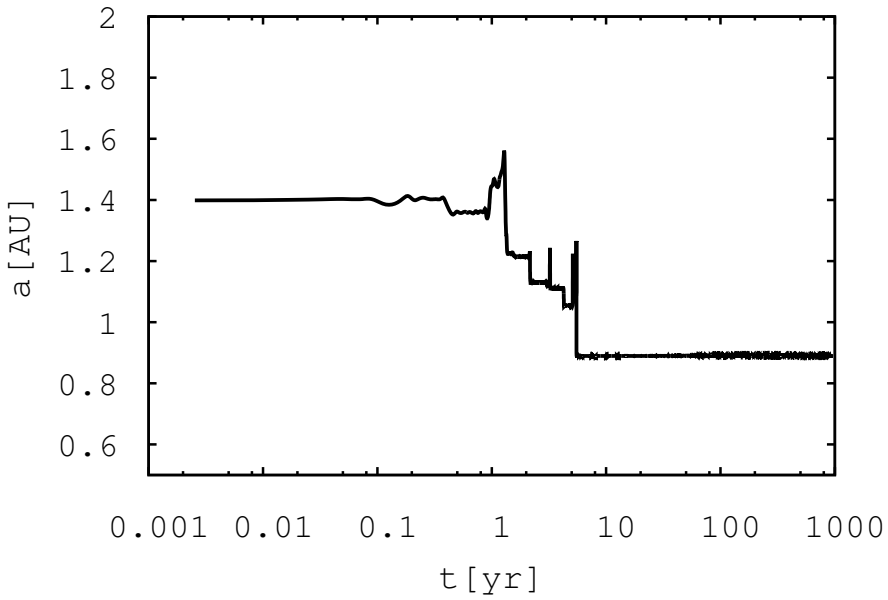

Figure 8: Evolution of the semi-major axis of a typical binary in a system. The sharp variations are due to stellar scatterings, which mostly result in a hardening of the binary. After the multiple system gets stabilized and isolated (which happens after ≈ 10 years), the decrease rate of the semi-major axis depends on dynamical friction alone.

Small scale	n_6	n_4
runs	86	86
$\langle a_{t=500\mathrm{yr}} \rangle$ [AU]	1.37	1.42
τ_{df}[yr]	$\sim 10^{13}$	$\sim 10^{15}$
Companion stars	3rd massive star, $11 \sim 12$ $\mathrm{M}_\odot$	
P_{HMXBc}	0.070	0.070
F_{HMXB} [10^{-4} $\mathrm{M}_\odot^{-1}$]	4.6	4.6

Table 5: Summary of results for simulations of 5-body groups forming on small scales. $\langle a_{t=5000\ \mathrm{yr}} \rangle$ indicates the semi-major axis at $t = 5000$ yr, while τ_{df} represents the dynamical friction timescale required for $a_{t=5000\ \mathrm{yr}}$ to shrink to a_{RL} (see equation 43). P_{HMXBc} is the fraction of runs in which a HMXBc forms, and F_{HMXB} is the number of HMXBc formed across all simulations, normalized by the total mass of the stars in the simulations.

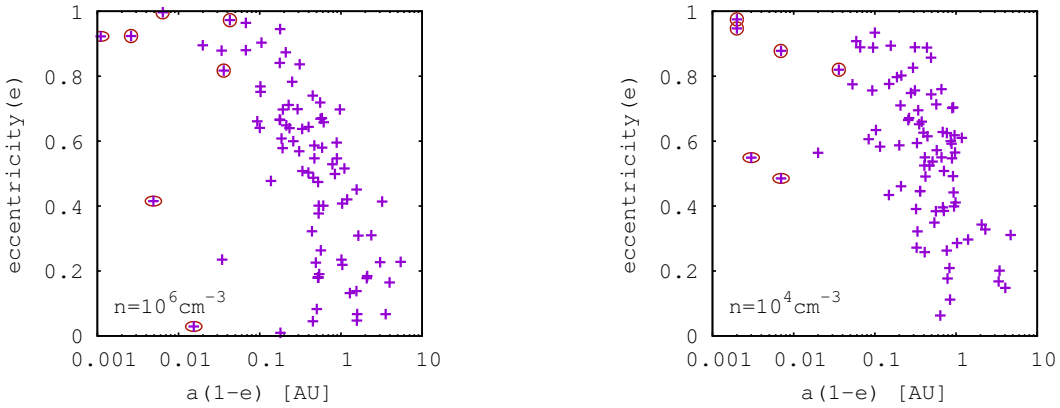

Figure 9: *Left*: Pericenter distance - eccentricity distribution plot for n_6. Two HMXB candidates (HMXBc) have been produced. *Right*: The same distribution plot for n_4. The circled points indicate HMXBc whose mass transfer via RLOF may occur periodically due to their eccentric orbits. The elliptically-circled point indicates an HMXBc whose semimajor axis is smaller than a_{RL}, so the mass transfer will be steady. There are a couple of binaries with pericenter distance smaller than R_{RL}, but they are excluded because $M_2 < 8$ $\mathrm{M}_\odot$.

and unbound single stars (in 15 out of 20 cases), in the small-scale runs a multiple system such as triple or quartet (rather than a simple binary) forms (in 59 of 86 cases) Less massive stars take some energy from the multiple system and convert it into their kinetic energies while causing the binary to harden.

As an example, Figure 8 shows the evolution in the semi-major axis of a compact binary in one of the runs. One can easily notice the quick decrease in the semi-major axis during violent stellar interactions (before 10 years), which clearly indicates the hardening. After a couple of stars are cast away and the multiple system is stabilized and isolated (after about 10 years), dynamical friction plays the main role in the decrease of the semi-major axis; after this point, we do not expect further dramatic hardening of the binary. In our 86 runs, the last surviving binary is typically of type B_{13} (pairing of the most massive and third most massive stars). with total mass of 130 $M_\odot$ and $\langle M_{S_3} \rangle = 11$ $M_\odot$. $\langle a_{t=5000\ \mathrm{yr}} \rangle = 1.87$ AU and the minimum semi-major axis is 0.2 AU. The corresponding dynamical friction time scale (equation 43) is $\sim 10^{13}$ yr. The ejected stars have speeds of $(10 - 100)$ c_s , and $c_s \sim 4$ km/s. Six of the HMXB candidates (HMXBc hereafter) have been formed in all runs ($P_{\mathrm{HMXBc}} = 0.070$). For later use, let us define F_{HMXB} as the number of HMXBc formed across all simulations, normalized by the total mass of the stars. Then $F_{\mathrm{HMXBc}} = 4.6 \times 10^{-4}$ $M_\odot^{-1}$ (This term will be used to estimate the X-ray luminosity and is one of the primary results of our study).

The outcomes of the n_4 simulations are quite similar: a triple or higher multiple forms in 65 of 86 runs, and $\langle a_{t=5000\ \mathrm{yr}} \rangle = 1.42$ AU. Furthermore, due to the same number of HMXBc, P_{HMXBc} and F_{HMXBc} are the same. One notable difference is that for n_4, the dynamical friction time scale is longer by 2 orders of magnitude compared to n_6, because this quantity is inversely proportional to the number density.

Interestingly, there are four runs (for each density value) in which the most compact binary is eccentric, and inside the requisite separation for RLOF at pericenter but outside it at apocenter. We consider only two of them as HMXBc and rule out the other two binaries since the mass of the more massive star is smaller than 8 $M_\odot$ (Heger et al., 2003). We present the distribution of eccentricity for pericenter distance in Figure 9. In particular, the *left* panel shows the distribution for n_6 calculations and the *right* panel for n_4. The circled point indicates HMXBc with distance at pericenter shorter than the corresponding Roche-Lobe radius.

How accretion proceeds in a highly eccentric binary system under these conditions remains an unsettled issue to date, as studies have claimed that the orbital semimajor axis and eccentricity can either increase or decrease depending on the binary properties at pericenter (Sepinsky et al., 2007, 2009). If the RLOF does induce circularization, then accretion proceeds normally (i.e. steadily). However, if the RLOF instead increases the eccentricity of the system, then, whether accretion can proceed steadily rather than intermittently will depend on the relative timescale between the disk lifetime τ_{disk} (on the order of the viscous timescale), and the orbital period of the binary t_{orb}. We computed these timescales for all the eccentric binaries in our simulations (for which the Roche-Lobe radius straddles the pericenter and apocenter), and found that $\tau_{\mathrm{disk}} > t_{\mathrm{orb}}$ in all cases but one.

This implies that the fraction of binaries whose eccentricities cause intermittent RLOF is small, and that as a global average, RLOF is steady to a good approximation.

The magnitude of the gas density considerably influences the characteristics of the dy-

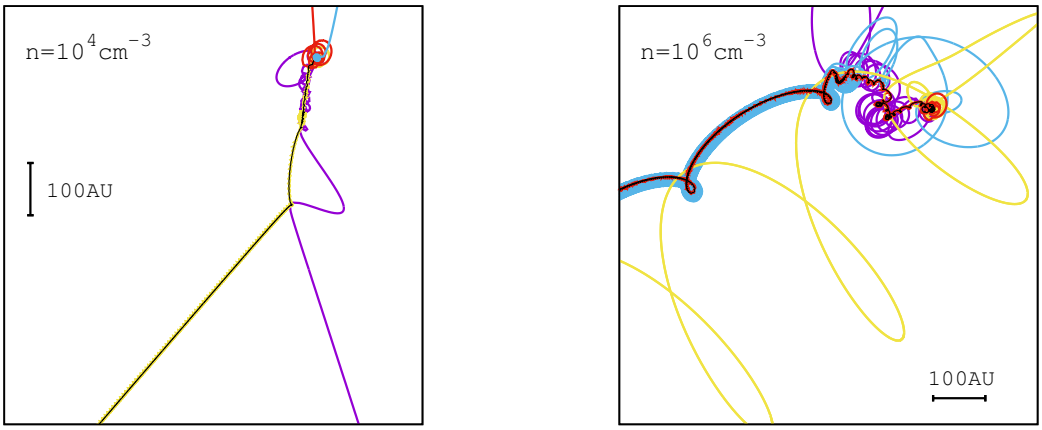

Figure 10: Trajectories up to $t = 2000$ yr for 5-body simulations with identical initial conditions but different densities: n_4 (*left* panel) and n_6 (*right* panel). Despite having the same starting point, the trajectories of the 5 stars evolve quite distinctly in the two backgrounds due to the different magnitudes of dynamical friction. The two simulations for different bound systems: a binary for n_4 (black+yellow line) and a quartet for n_6 (black+red+blue+yellow lines).

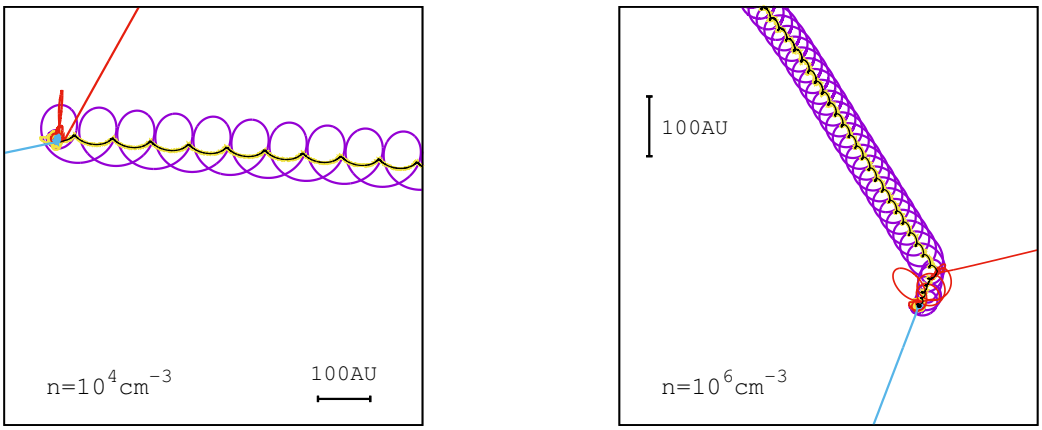

Figure 11: Trajectories of two 5-body simulations with identical initial conditions but different ambient gas densities. The simulations have formed triples with the identical stars (black+yellow+purple lines; contrast with 10) at $t = 1000$ yr, despite differences in the trajectories of each star and the center of mass.

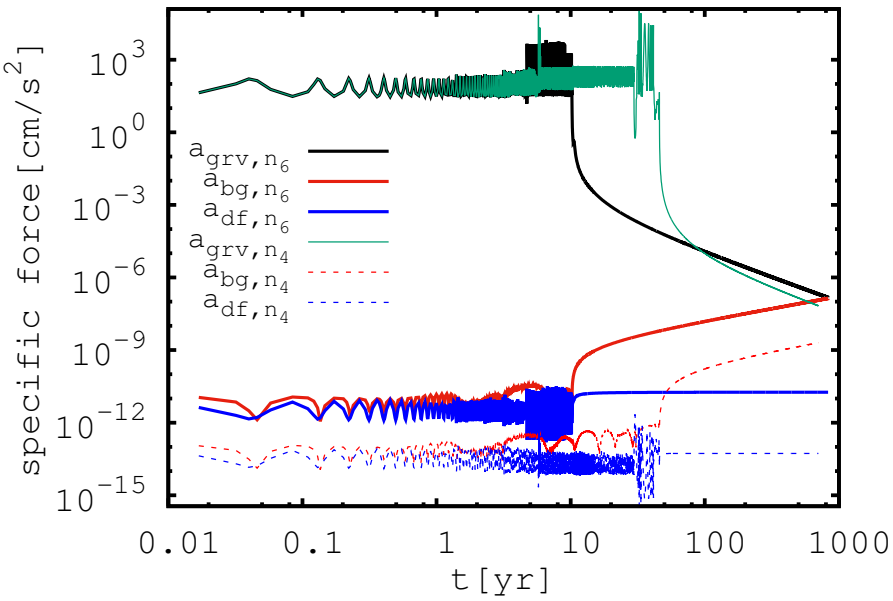

Figure 12: Changes with time in the three forces per unit mass (gravity from other stars, dynamical friction and background gravity) acting on star$_4$ (the light blue color line in Figure 11). We run two simulations with identical initial conditions but different ambient gas densities n_4 and n_6. During the early phases of the interaction, the gravitational force ($a_{\rm gr}$) dominates by several orders of magnitude over dynamical friction ($a_{\rm df}$) and the background force ($a_{\rm bg}$). Note that $a_{{\rm gr},n_6}$ (black line) and $a_{{\rm gr},n_4}$ (green line) synchronize at early times because the motion of the star is very close to Keplerian and $r \sim 1/v^2$. After the star is ejected at 10 yr for n_6 (later for n_4), $a_{\rm gr}$ monotonically decreases and $a_{\rm bg}$ increases ($a_{\rm bg} \sim r$ where r is the distance from the center of the halo). During the same time, $a_{\rm df}$ barely changes since the star moves in the supersonic regime and the speed decreases very slowly.

namical interactions. While on the one hand 40% of the simulations end up forming the same triples for both density values, on the other hand their trajectories and the center of mass movements differ significantly. Figure 10 shows two sample trajectories of 5 stars with low number density (left panel) and high number density (right panel) after 1000 yr. They were given identical initial conditions for the run, but their trajectories have developed differently. In Figure 11, at 800 yr, even though the same stars form a triple and the same star is ejected for both number densities, their trajectories are clearly different.

As can be seen in Figure 12, which shows the change of forces per unit mass exerted on a typical kicked-off star, $a_{\rm df}$ and $a_{\rm bg}$ are negligible compared to $a_{\rm gr}$ before about 5 years.

Note that $a_{\rm df}$ and $a_{\rm bg}$ are synchronized at early times because the motions of the stars are close to the Keplerian motion, $r \sim 1/v^2$. The star in Figure 12 is ejected at $t \approx 10$ yr for the high number density (and at $60 \lesssim t \lesssim 80$ yr for the low number density). This can be understood from the fact that $a_{\rm gr}$ monotonically decreases and $a_{\rm bg}$ increases ($a_{\rm bg} \sim r$ where r is the distance from the center of mass). At the same time, $a_{\rm df}$ barely changes since the star moves in the supersonic regime and the speed decreases slowly.

There is no noticeable difference in the overall results between the n_4 and n_6 runs—quantities such as $\langle a_{t=5000\ {\rm yr}} \rangle$, the dynamical friction time scale $\tau_{\rm df}$, the total mass of the most compact binary (or which star forms the compact binary with S_1) as summarized in Table 5. For both number densities, the companion star of the binary is typically the third most massive star S_3.

To sum up, we find that $P_{\rm HMXBc} \sim 7\%$ of our simulations form HMXBc, regardless of the gas density value. Normalized to the total stellar mass in the simulations, the number of HMXBc formed per stellar mass is $F_{\rm HMXB} \approx 4.6 \times 10^{-4}\ {\rm M}_\odot^{-1}$.

3.3.3 10-body simulations

We now explore several different configurations for the star group, and run several sets of simulations with 10 stars (instead of 5) with $n = 10^6\ {\rm cm}^{-3}$. These are: (1) 10-body version of the small scale calculation presented above; (2) head-on crash of two star groups containing 5 stars each; (3) a close encounter and subsequent inspiral and merger of two star groups containing 5 stars each. The latter two scenarios are motivated by the fact that the merger timescales and mass accretion timescales of Pop III host haloes, as well as the lifetimes of the massive Pop III stars themselves, are of the same order, ~ 10 Myr. This suggests that merging haloes will be continuously forming new stars (perhaps Pop II instead of Pop III) as they merge with other haloes, and that close interactions and mergers of nascent star groups may be relatively common. We generate stellar masses in the same way as for the 5-body case, but with a larger value for the parameter $M_{\rm max} = 300\ {\rm M}_\odot$. We have run each simulation for 500 yr. The results of these simulations are summarized in Table 6. We briefly discuss each one, as follows.

1. Scenario 1: 10-body group in isolation

   We set up the simulations as in the 5-body calculations, but with 10 stars.

   We find that $\langle a_{t=500{\rm yr}} \rangle = 1.0$ AU. Interestingly, there are 18 out of 54 cases in which $a < a_{\rm RL}$. There is a large difference in scale between $a_{\rm B_{12}}$ and $a_{\rm rest}$, where $a_{\rm B_{12}}$ is the semi-major axis of B_{12} (B_{12} binary made up of the two most massive stars) and $a_{\rm rest}$

Scenario	1	2	3	4A	4B
runs	54	30	30	30	30
$\langle a_{t=500\,\mathrm{yr}} \rangle$ [AU]	1.0	1.1	0.90	1.8	1.6
$\langle a_{\mathrm{B}_{12}} \rangle$ [AU]	1.6	1.5	1.8	2.4	2.0
$\langle a_{\mathrm{rest}} \rangle$ [AU]	0.72	0.15	0.42	1.2	0.23
P_{HMXBc}	0.33	0.33	0.27	0.13	0.27
F_{HMXBc} [10^{-4} $\mathrm{M}_\odot^{-1}$]	15	11	9.0	4.2	8.4

Table 6: Summary of the results from 10-body simulations. We have considered five different scenarios. Scenario 1 simulated the 10-body version of the 5-body calculation, i.e. an isolated star group. The other scenarios all involve collisions of two 5-body star groups. Scenario 2 is a head-on collision of two coplanar, co-rotating star groups. Scenario 3 also collided two groups of stars with co-rotating orbital planes, but with an impact parameter comparable to the sizes of the groups, which results in an inspiral and eventual merger. Scenarios 4A and 4B are similar to scenario 3, except that the orbital planes of the colliding star groups had mutual inclinations of 45° and 135°, respectively.

is the semi-major axis of the binary stars other than S_1 and S_2. This is a common feature of the 10-body simulations: they often end up with triples whose inner binary is B_{rest} while the outer binary is B_{12}. Our simulations yield $\langle a_{\mathrm{rest}} \rangle = 0.72$ AU and $\langle a_{\mathrm{B}_{12}} \rangle = 1.6$ AU. Also note that, in 14 out of 18 HMXBc, the binary is $\mathrm{B}_{\mathrm{rest}}$ (i.e. it is not made up of the two most massive stars).

Since the compact binaries in these simulations form quickly and we only follow them for 500 years, it is technically possible that they will be disrupted before one of the stars turns into a CO. However, our simulations for the 5-body scenario showed that compact, quasi-steady binaries are unlikely to be disrupted, and for practical purposes we extrapolate this qualitative result to the 10-body case.

We find that a HMXBc forms in a larger fraction of these simulations than in the 5-body case, $P_{\mathrm{HMXBc}} = 0.33$, for the obvious reason that there are more stars. Per unit stellar mass in the simulations, the number of HMXB candidates is $F_{\mathrm{HMXB}} = 1.5 \times 10^{-3}\ \mathrm{M}_\odot^{-1}$, which is a factor ≈ 3 higher than we found for the 5-body case.

2. Scenario 2: Collision between two 5-star groups – head-on collision

   Two groups of 5 stars are set up with random initial conditions, in the same manner as for the previous simulations of 5-body groups. The two groups are then arranged to collide head-on, as follows: they are placed at a separation of two to three times their sizes and their disks are aligned so that the mutual inclination is zero. The initial relative speed of the groups is roughly the speed of sound and the center of mass of one group is set to move directly toward the center of mass of the other group.

   We find that prior to colliding, each group forms a compact binary of type B_{12} (the most and second-most massive star). When the two groups collide, those two binaries that existed before the collision were broken and the two most massive stars of each group form a new compact binary with high chances.

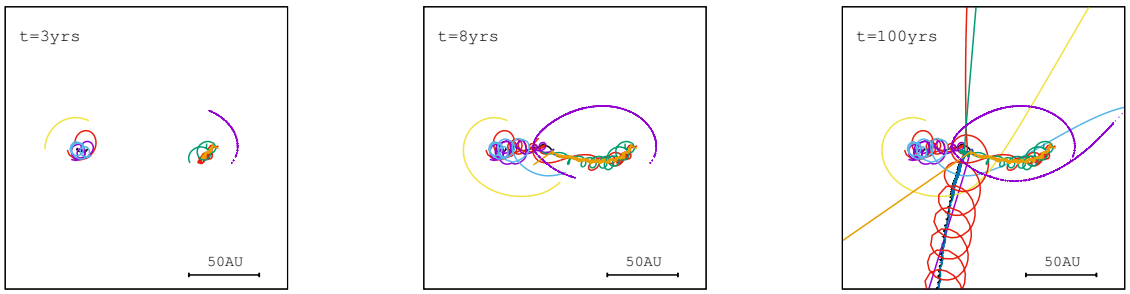

Figure 13: 10-body head-collision (scenario 2). In these sample trajectories, five stars in each group are in nearly Keplerian motion. The groups move towards each other with a relative velocity of $\sim c_{\rm s}$ (left panel). At $t \approx 8$ yr (middle panel), the two haloes begin to merge. During the merger, all ten stars undergo stellar scatterings (right panel). One triple (black+dark blue+red lines) has been formed and it is moving in -y direction. In these head-on collisions, it is likely for multiple systems that existed before merging to be broken, while a few new multiple systems are formed.

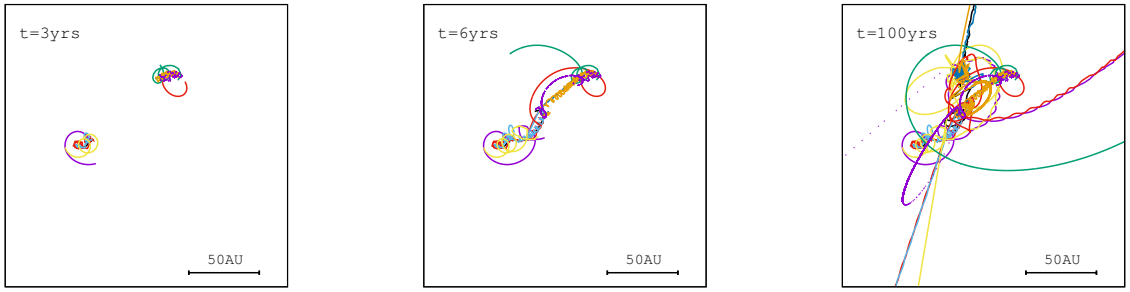

Figure 14: 10-body spirally merging case (scenario 3). As with scenario 2, two groups of 5 stars are set on a collision cours. However, in this scenario the impact parameter is $\sim 2-3$ times the size of groups, whereas in scenario 2 it is set to be zero. The groups are approaching each other at the relative velocity of $c_{\rm s}$ (left panel). At $t = 8$ yr, the groups are about to merge (middle panel). After some time (right panel), two binaries are ejected (red and purple lines in the x-direction and black and blue lines in the y-direction). In this sample case, the most compact binary is the one ejected in the y direction (black+blue lines) with $a = 0.8$ AU at $t = 100$ yr. A difference between this scenario and the head-on collision is that compact systems are more likely to survive the merger.

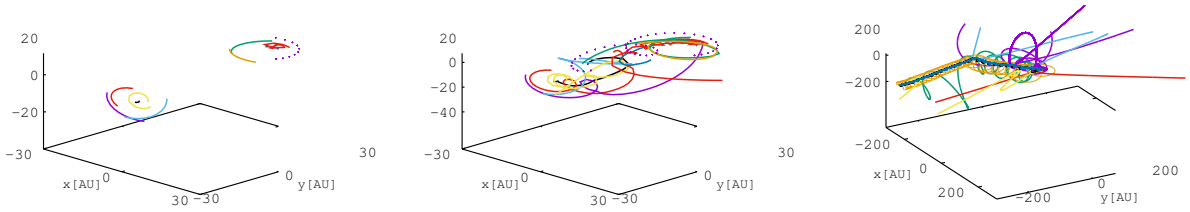

Figure 15: Trajectories for scenario 4A, a collision between two 5-body groups, each on quasi-Keplerian orbits, but with the orbital plane of the two groups tilted with respect to each other at an inclination $i = 45°$. The setup is the same as scenario 3, except for the mutual inclination of the orbital planes of the colliding star groups. The left panel shows the two groups on a collision course. At $t = 4$ yr, the halos are about to merge (middle panel). After $t = 10$ yr (right panel), a triple having the most compact binary are ejected [black and blue lines (inner binary) and brown line].

The average $\langle a_{t=500\ \mathrm{yr}} \rangle$ is 1.1 AU, but $\langle a_{\mathrm{B}_{12}} \rangle = 1.5$ AU and $\langle a_{\mathrm{rest}} \rangle = 0.15$ AU, meaning that the most compact binaries are not formed from the most massive stars. The shorter average separations may be a result of a larger number of early 3-body scatterings, which act to harden the group as a whole.

HMXBc form in 10 out of 30 runs, and they are not of type B_{12}. However, we find a rate of HMXB formation per stellar mass $F_{\mathrm{HMXB}} = 1.1 \times 10^{-3}\ \mathrm{M}_{\odot}^{-1}$; this is higher than in the 5-body case and comparable to Scenario 1 above. Sample trajectories for one of the simulations of scenario 2 are depicted in Figure 13.

3. Scenario 3: Collision between two 5-star groups – spirally merging case

   We have used the same input parameters for the two groups as in Scenario 2, except that we now set the impact parameter to be of order the size of the group, whereas it was set to zero in Scenario 2.

   The groups are given opposite velocities of $\sim c_{\mathrm{s}}$ along the x-direction, and are offset by a displacement along the y-direction that is $\sim 2 - 3$ times the typical size of the star group (~ 20 AU) so that they merge with a spiral motion.

   We find that $\langle a_{t=500\ \mathrm{yr}} \rangle = 0.90$ AU, $\langle a_{\mathrm{B}_{12}} \rangle = 1.8$ AU and $\langle a_{\mathrm{rest}} \rangle = 0.42$ AU.

   The average separation lies between what we find in Scenarios 1 and 2. This can be interpreted as being due to the fact that these simulations (in which the two groups merge gradually via inspiral) have more close 3-body interactions than in Scenario 1 (in which 10 stars in quasi-Keplerian orbits evolve in isolation) but fewer such interactions than in Scenario 2 (in which the two groups merge head-on).

   HMXBs form in 8 out of 30 runs and, as with Scenarios 1 and 2, none of the HMXBs are made up of the two most massive stars. We find a similar HMXB formation rate per stellar mass, $F_{\mathrm{HMXB}} = 9.0 \times 10^{-4}\ \mathrm{M}_{\odot}^{-1}$. Sample trajectories from one of the simulations for Scenario 3 are shown in Figure 14.

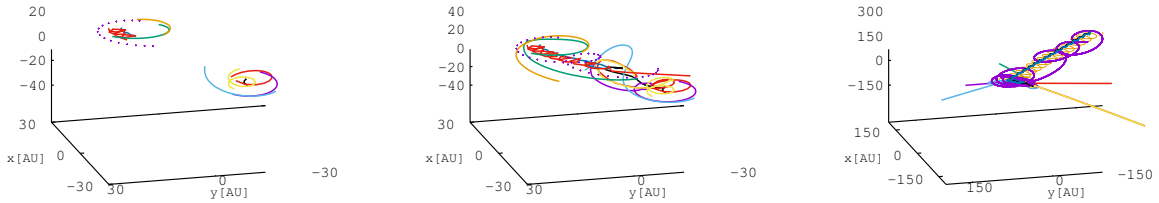

Figure 16: Trajectories for scenario 4B—the same as scenario 4A (Figure 15), but with the two groups nearly counter-rotating with respect to each other ($i = 135°$). At $t = 4$ yr, the groups are about to merge (middle panel) and after $t = 10$ yr (right panel), a quartet containing the most compact binary is ejected toward the upper right of the panel [black+blue lines (inner binary), brown and thick purple lines].

4. Scenarios 4A and 4B: Collision between two 5-body groups – spirally merging case with inclinations of 45 degrees and 135 degrees

   In these two scenarios, we again set two groups of five stars each on a collision course. The difference is that the orbital plane of one star group is tilted, so that the two stellar disks have a mutual inclination i. We set the inclination at $i = 45°$ (nearly co-rotating) for scenario 4A, and $i = 135°$ (nearly counter-rotating) for scenario 4B. The groups are initially placed at a separation of two to three times their sizes, and set in motion at the same speeds as for the inspiral case (scenario 3).

   We find that $\langle a_{t=500\,\mathrm{yr}} \rangle = 1.8$ AU, $\langle a_{\mathrm{B}_{12}} \rangle = 2.4$ AU, and $\langle a_{\mathrm{rest}} \rangle = 1.2$ AU for scenario 4A. In scenario 4B, we find $\langle a_{t=500\,\mathrm{yr}} \rangle = 1.6$ AU, $\langle a_{\mathrm{B}_{12}} \rangle = 2.0$ AU, and $\langle a_{\mathrm{rest}} \rangle = 0.23$ AU. In 70% of the runs for both scenarios, the two most massive stars form the most compact binary. Stellar binaries end up with somewhat closer separations in the nearly counter-rotating case, due to the fact that the net angular momentum of the merged star group is smaller. Indeed, we find a total of four HMXBc across all the $i = 45°$ simulations, and eight in a same number of $i = 135°$ simulations.

   For the same reason, we find a larger fraction of HMXB candidates per stellar mass simulated in the nearly counter-rotating case ($F_{\mathrm{HMXB}} = 8.4 \times 10^{-4}$ M$_\odot$) compared to the nearly co-rotating case ($F_{\mathrm{HMXB}} = 4.2 \times 10^{-4}$ M$_\odot$). The overall formation rate of HMXB candidates is lower for both cases than the cases in which all the stellar orbits were nearly coplanar (Scenarios 1, 2 and 3), plausibly due to the additional degree of freedom in the stellar orbits. Still, the value of F_{HMXB} is within a factor of a few for all of our simulations.

   Sample trajectories from runs for scenarios 4A and 4B are depicted in Figure 15 and Figure 16, respectively.

3.4 Discussion

3.4.1 Binary evolution and formation of HMXB candidates

Our large-scale simulations show that the if Pop III stars form hundreds of AU apart, as in the simulations of Stacy & Bromm (2013), then the timescales required to make HMXBs via stellar scatterings are simply too long. On the other hand, if protostellar clouds fragment and form stars in close groups on scales of $\gtrsim 10$ AU, as in Greif et al. (2011), then a small fraction of groups can form HMXBs. We briefly discuss the dynamics of HMXB formation in our simulations, then move on to discuss the astrophysical implications of our findings.

The simulations indicate that, as expected, scatterings play a major role in making a compact binary. The background potential and the dynamical friction play a secondary role, by allowing the most compact binary (or triple) to remain near the center of mass of the halo and for other stars to return and scatter again and again.

We find that on average, the number of HMXB candidates formed per stellar mass, $F_{\rm HMXB}$, is a function of the number and orientation of close 3-body encounters. Our results indicate that $F_{\rm HMXB}$ may be somewhat higher in configurations that result in fewer ejections of stars, and if the interactions are coplanar. While we are able to interpret this, as well as trends in the average separation between stellar pairs, in terms of the initial kinematic setup of the various scenarios simulated (see §3.3.3 above), the value of $F_{\rm HMXB}$ does not vary by more than a factor ≈ 3. We interpret this lack of a significant variation in $F_{\rm HMXB}$, for such a diverse set of initial conditions and ambient gas densities, to mean that our values are not far from the one that results from similar stellar encounters in nature.

3.4.2 The effect of migration on the formation of HMXBs

Another way in which a nascent stellar group could harden is migration through a gaseous disk. The migration could occur as the protostars form—Greif et al. (2012) found significant accretion from the protostellar disk onto the most massive protostar, and did not follow the evolution of the system beyond this stage. (It could also occur to a lesser degree in a vestigial gas disk, after the stars are in place.)

We evaluate the possible role of disk migration on the separation of Pop III stars by considering a steady, geometrically thin disk with an α viscosity (Shakura & Sunyaev 1973; see also Frank et al. 2002). We adopt a disk with $\alpha = 0.01$ and an accretion rate $\dot{m} \sim 10^{-3}$ M$_\odot$ yr^{-1}, following Tan & McKee (2004) and Tan & Blackman (2004), who considered the structure of accretion disks around Pop III stars at high redshifts.

We estimate the migration timescale $\tau_{\rm mig}$ following Syer & Clarke (1995). We take a binary system with primary mass $M_1 = 120$ M$_\odot$ and secondary mass $M_2 = 11$ M$_\odot$, based on the mean values found across our simulations. For these masses and disk parameters, the secondary is able to clear a gap around its orbital path (Syer & Clarke, 1995; Seager, 2010; Lubow & Ida, 2010), by satisfying both of the following two conditions:

$$(1)\ \frac{H}{R} \leq \left(\frac{q}{\alpha}\right)^{1/2}, \tag{48}$$

$$(2)\ \frac{H}{R} \leq \left(\frac{q^2}{\alpha}\right)^{1/5}. \tag{49}$$

where R is the distance of the secondary from the primary, H is the scale height of the disk at that location, and $q = M_2/M_1 \sim 0.1$ is the binary's mass ratio. If condition (1) is violated, the gap will be closed by the radial pressure gradient. Condition (2) relates the gap width and the Roche radius of the secondary; this ensures that the secondary acts to transfer orbital angular momentum through the disk, rather than accreting mass via RLOF. We find that inside $R \sim 15$ AU, for the disk properties stated above, H/R is smaller than the right-hand side of equation (48) by a factor $\gtrsim 80$, and smaller than the right-hand side of equation (49) by a factor $\gtrsim 8$, indicating that the secondary is easily able to open a gap.

The migration timescale of the secondary depends on the dimensionless parameter (Syer & Clarke, 1995)

$$B \equiv \frac{4\pi\Sigma_0 R^2}{M_2}, \tag{50}$$

where Σ_0 is the local surface density of a steady-state disk around the primary, in the absence of perturbations by the secondary. For $B > 1$, the gas in the disk is able to dynamically dominate over the gravitational influence of the secondary, and the secondary is pushed inward on the viscous diffusion timescale of the disk,

$$\tau_{\rm mig,0} \sim \alpha^{-1}\Big(\frac{H}{R}\Big)^{-2}\Omega^{-1}\,. \tag{51}$$

For $B < 1$, Syer & Clarke (1995) found that the migration timescale is longer,

$$\tau_{\rm mig,1} \sim \frac{1}{B^{7/17}}\,\tau_{\rm mig,0}\,. \tag{52}$$

In Figure 17, we plot the local disk mass near the secondary, $4\pi\Sigma_0 R^2$, alongside the typical secondary mass, 11 $M_\odot$ (the numerator and denominator, respectively, for the ratio B). For $R \lesssim 15$ AU, $B < 1$ and the migration timescale is expected to slow. We plot the two migration timescales $\tau_{\rm mig,0}$ and $\tau_{\rm mig,1}$ in Figure 18.

Both of these timescales are shorter than both the typical lifetimes of protostellar disks, as well as the lifes of the stars themselves[7]. This suggests, along with findings that the disks may be able to survive even under radiation feedback (Stacy et al. 2012; Hosokawa et al. 2015), that disk migration could lead to initial stellar separations smaller than what we have assumed, making the formation of HMXBs via stellar scatterings more favorable.

We submit that the HMXB formation rates inferred from our simulations be taken as a conservative estimate, with possible additional contributions from channels other than stellar scattering.

3.4.3 X-ray output

As discussed in §3.1, HMXBs are believed to be a major source of X-rays in the early universe. Observations of nearby star-forming galaxies suggest that their X-ray luminosities (which are dominated by HMXBs) scale linearly with their star formation rate (Grimm et al., 2003; Gilfanov et al., 2004; Persic et al., 2004; Mirabel et al., 2011). Mineo et al. (2012)

[7]Note that, if the adopted value of α were higher, e.g. $\sim 0.1 - 1$ as suggested by recent simulations of Pop III protostellar disks (Clark et al., 2011), the migration timescale would become even shorter.

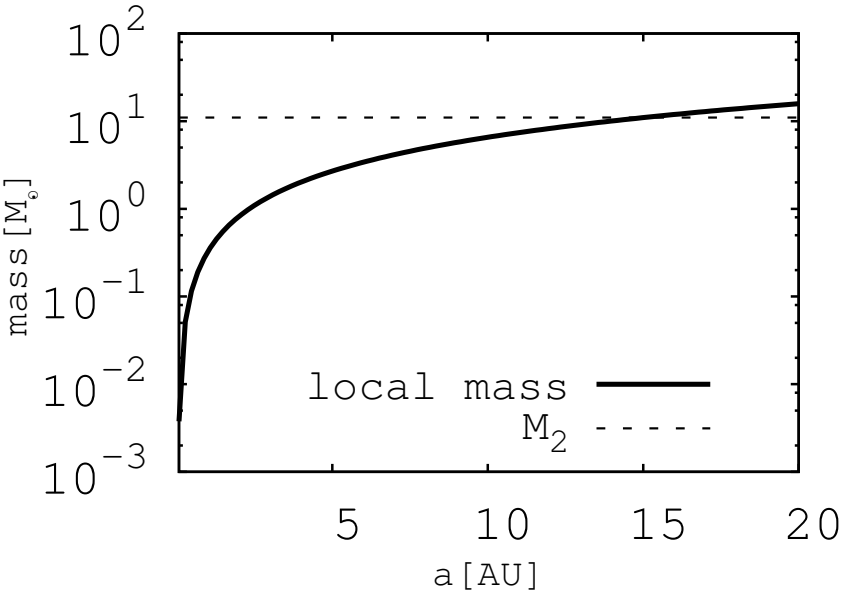

Figure 17: A comparison between the mass of a typical secondary of a binary in our simulations ($M_2 = 11$ M$_\odot$, dashed line) and the local mass $4\pi\Sigma_0 R^2$ (solid line) in a protoplanetary disk near its orbit. The migration timescale increases if the ratio $B = 4\pi\Sigma_0 R^2/M_2 < 1$, which is the case for close orbits.

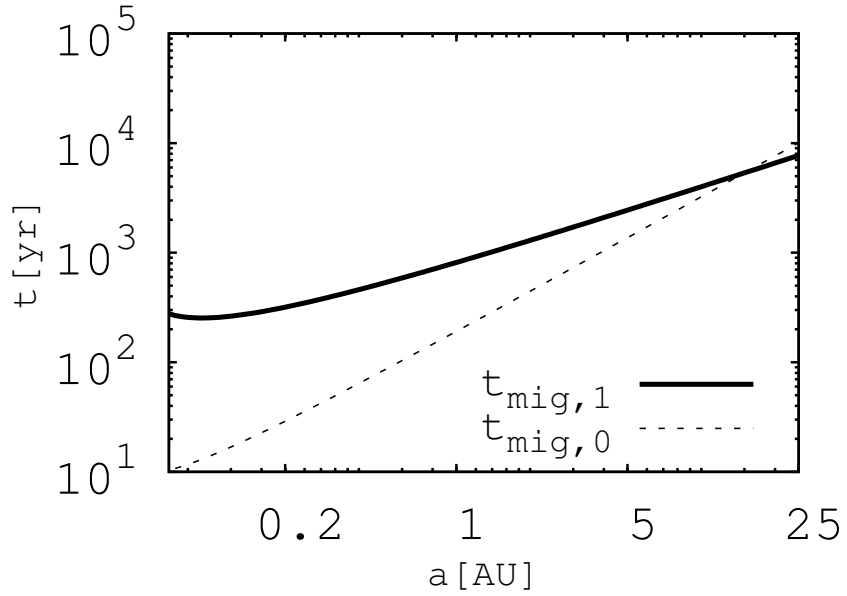

Figure 18: Migration timescale for a circumbinary protostellar disk, based on the typical binary properties of our simulations. The unperturbed Type II migration timescale, $\tau_{\rm mig,0}$ (dashed line) shows the timescale assuming that the secondary mass is much smaller than the local disk mass. A longer timescale $\tau_{\rm mig,1}$ (solid line) is expected if the secondary mass is large compared to the disk mass (see previous Figure).

find many studies support a linear proportionality between the X-ray luminosity of HMXBs and the star formation rate (SFR) (Grimm et al., 2003; Gilfanov et al., 2004; Persic et al., 2004; Mirabel et al., 2011)). In the linear regime, the X-ray luminosity of the local universe is given by Mineo et al. (2012),

$$L_{2-10\ \mathrm{keV}}^{\mathrm{local}} = 3 \times 10^{39} \times \frac{\mathrm{SFR}}{\mathrm{M}_{\odot}\ \mathrm{yr}^{-1}}\ [\ \mathrm{erg\ s}^{-1}], \tag{53}$$

where SFR is the star formation rate.

How the ratio of X-ray luminosity to SFR evolves with redshift is a key question in evaluating the properties of galaxies and young stellar populations, and in linking the earliest galaxies (most of which should be actively forming stars, based on their mass accretion rates) with their X-ray luminosities. The question is as yet unresolved by current observations (Dijkstra et al., 2012; Basu-Zych et al., 2013), and is often treated as a free parameter in studies estimating the X-ray production of the first galaxies (Furlanetto, 2006; Fialkov et al., 2014; Tanaka et al., 2016).

In the following, based on the results of our simulations, we are going to quantitatively evaluate the relation between X-ray luminosity and SFR, and compare it with equation (53). We can write the X-ray luminosity as

$$\begin{aligned} L_{2-10\ \mathrm{keV}} = L_{\mathrm{Edd}} \times f_{\mathrm{Edd}} \times f_{2-10\ \mathrm{keV}} \\ \times t_{\mathrm{acc}} \times f_{\mathrm{sur}} \times f_{\mathrm{esc}} \times F_{\mathrm{HMXB}} \times \mathrm{SFR}\,. \end{aligned} \tag{54}$$

Below, we discuss each quantity in equation (54).

1. L_{Edd}, the Eddington luminosity, which scales with the typical mass of the CO engine M_{CO} as $1.3 \times 10^{38} (M_{\mathrm{CO}}/\ \mathrm{M}_{\odot})$ erg s^{-1}. Note that, for stars which leave behind a BH remnant ($M \gtrsim 25 M_{\odot}$), we use the stellar mass as a proxy for the BH mass for simplicity, due to the theoretical uncertainties in evaluating the mass loss due to winds and during the transition to a BH. Any significant mass loss (e.g. simulations by Zhang et al. 2008 suggest that the BH remnants of massive metal-free stars end up with $\sim 40\%$ of the progenitor mass) would be directly translatable to the estimated X-ray luminosities reported in Table 7. For stars with $M \lesssim 25 M_{\odot}$, we consider two limiting cases: one in which all the stars form BHs, and one in which all the stars form NSs, for which we assume a typical mass value of 1.4 $\mathrm{M}_{\odot}$ (Lattimer & Prakash, 2005). The simulations by Zhang et al. (2008) for massive metal-free stars show that, depending on the explosion energy (and hence the amount of fallback) the compact remnant can be either an NS or a BH. However, this uncertainty plays a very small role for our results due to the fact that the number of HMXBs with a star of $M < 25$ $\mathrm{M}_{\odot}$ is a small fraction of the total. In particular, we find that there is only one star with $M < 25$ $\mathrm{M}_{\odot}$ in a HMXB for each of the 5-body simulations (n_6 and n_4) and 4 stars for *scenario 1*, and none for the other cases. The total X-ray luminosities, when assuming an NS remnant for any star with $M < 25$ $\mathrm{M}_{\odot}$, are smaller by at most a factor of 4% with respect to the values given in Table 7, calculated assuming a BH remnant for all the stars.

2. $f_{\rm Edd}$, the typical ratio of the total radiative power emitted by HMXBs (the bolometric luminosity) to $L_{\rm Edd}$. If the typical luminosity of a HMXB during an active phase is Eddington, then $f_{\rm Edd}$ is effectively the mean duty cycle. Other studies have adopted values ranging from 0.1 to 0.5 (Mirabel et al., 2011; Belczynski et al., 2008; Salvaterra et al., 2012); we take as our fiducial value $f_{\rm Edd} = 0.1$.

3. $f_{2-10\ \rm keV}$, the fraction of the bolometric luminosity that is emitted between 2 and 10 keV. Observational estimates vary between 0.1 and 0.8 (Sipior et al., 2003; Migliari & Fender, 2006). Because the BH masses in Pop III HMXBs are expected to be higher than in present-day populations, and the peak energies of accretion disks scale with the mass of the central engine as $M_{\rm BH}^{-1/4}$, their characteristic spectra could be somewhat softer. We do not expect this to significantly affect our estimates here.

4. $t_{\rm acc}$, the time that a massive binary spends as a HMXB, with the less massive star donating mass to the CO companion. If the two stars form simultaneously and form a compact binary before the more massive member dies to become a CO, then this is simply $t_2 - t_1$, the difference in the lifetimes of the stars. We use this (somewhat arbitrary) estimate. However, because the lifetimes of the less massive star ($t_2 \gtrsim 10$ Myr for stars with masses $\lesssim 10$ M$_\odot$) are comparable to the expected specific star formation rate in galaxies at this redshift, we argue that any prescription that satisfies $t_{\rm acc} \lesssim t_2$ is a reasonable order-of-magnitude estimate.

5. $f_{\rm sur}$, the fraction of HMXB candidates identified in our simulations that actually survive to become HMXBs. This quantity accounts for possible disruptions of binaries, due to (a) the merger of the stars during main sequence and post-main sequence evolution (Power et al., 2009); (b) the more massive star getting kicked following a supernova explosion (Repetto et al., 2012; Janka, 2013); and (c) subsequent disruptions by stellar scatterings that were not captured by our simulations. (Our simulations follow the evolution of star groups until the formation of a stable compact binary, but not until stellar death millions of years later.) Theoretical estimates typically yield $\sim 0.2 - 0.3$ (Jeon et al., 2014; Artale et al., 2015)

6. $f_{\rm esc}$, the fraction of X-rays that escape the galaxy. Unless the environment of the HMXBs are Compton-thick, which is unlikely for the low-mass galaxies of interest, we expect $f_{\rm esc} \lesssim 1$.

7. $F_{\rm HMXB}$, the number of HMXBs formed per stellar mass. This is the main output of our simulations. (Note that $F_{\rm HMXB}$ has units $\rm M_\odot^{-1}$; we use the capital letter to distinguish it from the dimensionless fractions represented by f including the mass fractions such as $f_{\rm HMXB}$ from Hummel et al. (2015).) Whereas previous theoretical works had arrived at this value by extrapolating the locally observed value with an assumed redshift evolution, or with free parameters, here we provide an estimate based on suites of N-body simulations whose initial conditions are motivated by cosmological simulations of Pop III star formation. For example, Hummel et al. (2015) estimated the early X-ray background and its effect on baryonic infall into minihalos. Those authors arrived at a value of $f_{\rm HMXB}$ that is a factor of a few smaller than ours. This discrepancy is likely due

to two factors. First, they adopted a flat IMF, whereas our top-heavy IMF produces more massive stars, $\gtrsim 0.7$ of which turn into BHs. Second, because more massive binaries are more likely to survive three-body interactions, the compact binaries at the end of our simulations have an even larger fraction of their masses in BHs. A higher mass fraction in BHs should result in more X-ray production per unit star formation, and a more pronounced effect on the IGM kinetic temperature and greater thermal feedback on early galaxy formation and evolution (cf. Ripamonti et al. 2008; Tanaka et al. 2012a).

Across all of our small-scale simulations—varying the number of stars in the group, whether groups evolved in isolation or through several different orientations of mutual collisions, and exploring two values for the ambient gas density separated by two orders of magnitude—we find $F_{\rm HMXB} \sim 10^{-3}$, varying by less than a factor of 4 between the lowest and the highest values (see Table 7).

Finally, we can write $L_{2-10\ \rm keV}$ as follows :

$$\frac{L_{2-10\ \rm keV}}{\rm SFR} = 0.33 \times \frac{M_{\rm BH}}{\rm M_\odot} \times \frac{f_{\rm Edd}}{0.1} \times \frac{f_{2-10\ \rm keV}}{0.1} \times \frac{t_{\rm acc}}{\rm Myr} \times \frac{f_{\rm sur}}{0.5} \times \frac{f_{\rm esc}}{0.5} \times \frac{F_{\rm HMXB}}{10^{-3}\ \rm M_\odot^{-1}} \times \frac{10^{39}\ \rm erg\ s^{-1}}{\rm M_\odot\ yr^{-1}}\,. \quad (55)$$

Based on our choices of the factors, $f_{\rm Edd} = 0.1$, $f_{2-10\ \rm keV} = 0.1$, $f_{\rm sur} = 0.5$ and $f_{\rm esc} = 0.5$, we can estimate the normalized X-ray luminosities per SFR, $L_{2-10\rm keV}$/SFR. We report this quantity for each of our models in Table 7. It varies from a minimum of 37 to a maximum of 450 among the studied scenarios.

These $L_{\rm X}$-to-SFR ratios are $\sim 40-150$ higher than what is observed in the local Universe. This result is qualitatively consistent with the findings of Basu-Zych et al. (2013) and Kaaret (2014), who find an increase in the $L_{\rm X}$-to-SFR ratios toward $z \gtrsim 4$. Our high $L_{\rm X}$-to-SFR values stem from the large mass of the HMXBc primary, and the relatively low mass of the secondary. The former leads to a higher Eddington luminosity compared to typical stellar-mass BHs ($\sim 3\ \rm M_\odot$) in the local Universe, and the latter results in long stellar lifetimes, which in turn leads to longer $t_{\rm acc}$.

3.4.4 Implications for the thermal history of the IGM and the 21cm radiation

A higher $L_{\rm X}$-to-SFR ratio implies that IGM heating will occur earlier than commonly thought. The thermal history of the IGM can be probed in the 21 cm line, which is observable in absorption (or in emission), depending on whether the spin temperature of the IGM is below (or above) the CMB temperature.

If the IGM heats early, as suggested by our estimates of the X-ray emission of early galaxies, the 21 cm absorption line appears earlier, and the "dip" as a function of redshift caused by adiabatic cooling is not as deep and not as sharp as in the case of late heating as it would otherwise (see Figure 2 in Fialkov & Barkana 2014). Another consequence of early,

intense heating is that the temperature of the IGM could become high enough, to suppress the formation of low-mass galaxies (Ripamonti et al., 2008) and the growth of their nuclear BHs (Tanaka et al., 2012a).

3.4.5 Implications for Gamma Ray Bursts from Pop III stars

As discussed in §3.1, the fraction of HMXBs at high redshifts has potential implications for the expected rates of LGRBs from Pop III stars. According to the collapsar model (MacFadyen & Woosley, 1999), for an exploding massive star to yield a GRB, several conditions need to be satisfied, namely:

1. The core of the star must collapse to a BH. This is realized by most Pop III stars, given their large masses.

2. The hydrogen envelope of the progenitor star must be stripped, so that a relativistic jet can penetrate and exit the remaining envelope.

3. The BH should be surrounded by an accretion disk of high angular momentum material. This is realized if the core of the progenitor star has retained sufficient angular momentum during the evolution.

Binary systems more easily satisfy the last two conditions with respect to single stars (see e.g. Cantiello et al., 2007). In fact, for single stars to end their lives as LGRBs, they need to be born with large initial rotation (since they are less likely to be spun up by other stars), and also avoid being slowed down by magnetic torques (Spruit, 2002; Yoon et al., 2006; Perna et al., 2014).

In contrast, binary stars can spin up the helium core of the progenitor star via tidal coupling and spin-orbit locking. Further, RLOF can strip the hydrogen envelope during a common-envelope phase without reducing the rotation of the helium core (Bromm & Loeb, 2006). This is especially important for Pop III stars, whose heavier hydrogen envelopes would be more difficult to shed in isolation. Therefore, compact binary systems, or HMXBs, constitute a promising channel to produce LGRBs from Pop III stars.

We can use our results for the formation rates of Pop III HMXBs to estimate the fraction of LGRBs from Pop III stars[8]. Bromm & Loeb (2006) quantified the GRB formation efficiency as

[8]Note that in our study, since we are considering HMXB 'candidates', there is also the possibility of merger of the two stars during the common-envelope inspiral phase, when they both are stripped down to their Helium cores. This event could provide another avenue for the formation of GRBs (Fryer & Heger, 2005). Alternatively, after both stars have undergone the SN explosion, if the system is still bound, the compact objects of the binary, upon merger as a result of gravitational energy loss, would be likely contributors to the population of Short Gamma-Ray Bursts (e.g. Narayan et al. 1992).

	5-body (n_6)	5-body (n_4)	*Sce. 1*	*Sce. 2*	*Sce. 3*	*Sce. 4A*	*Sce. 4B*
$M_1[\,\mathrm{M}_\odot]$ ($\tau_{\mathrm{life},1}[\,\mathrm{Myr}]$)	110 (2.5)	88 (2.9)	110 (2.5)	120 (2.4)	110 (2.4)	160 (2.0)	110 (2.4)
$M_2[\,\mathrm{M}_\odot]$ ($\tau_{\mathrm{life},2}$[Myr])	12 (17)	11 (18)	18 (11)	45 (4.8)	28 (7.1)	63 (3.7)	63 (3.7)
t_{acc} $(=\tau_{\mathrm{life},2}-\tau_{\mathrm{life},1})$[Myr]	14	15	8.6	2.5	4.7	1.7	1.3
F_{HMXB} $[10^{-4}\,\mathrm{M}_\odot^{-1}]$	4.6	4.6	15	11	9.0	4.2	8.4
$L_{2-10\mathrm{keV}}/\mathrm{SFR}$ $[10^{39}\mathrm{ergs}^{-1}\,\mathrm{M}_\odot^{-1}\,\mathrm{yr}]$	220	200	450	110	160	37	40
$L_{2-10\mathrm{keV}}/L^{\mathrm{local}}_{2-10\mathrm{keV}}$	75	67	150	36	52	12	13
$\frac{\eta_{\mathrm{GRB,PopIII}}}{\eta_{\mathrm{GRB,PopI/II}}}$	2.2	2.2	7.0	5.4	4.3	2.0	4.0

Table 7: Summary of the results of X-ray luminosity and GRB efficiencies. In the table, second and third columns correspond to 5-body calculations and the rest columns (from *Sce. 1* to *Sce. 4B*) are the results from 10-body calculations with the number density of $10^6\mathrm{cm}^{-3}$.According to the ratio $L_{2-10\ \mathrm{keV}}/L^{\mathrm{local}}_{2-10\ \mathrm{keV}}$, where $L^{\mathrm{local}}_{2-10\ \mathrm{keV}}/\mathrm{SFR} = 3\times10^{39}$ erg s^{-1}, it is expected that the 2-10 keV X-ray luminosity is $\sim 10^2$ times larger than the X-ray luminosity of the local universe. These larger values essentially result from the large mass of the star 1. In the bottom row, we have listed the LGRB efficiency, derived from each set of simulations. The higher HMXB formation rates, lead to larger efficiencies than the estimated value for Pop I/II stars, $\eta_{\mathrm{GRB,PopI/II}} \simeq 4.2\times10^{-6}$ with beaming factor of 1/50 (e.g. Bromm & Loeb 2006).

$$\eta_{\rm GRB} \simeq \eta_{\rm BH}\,\eta_{\rm bin}\,\eta_{\rm close}\,\eta_{\rm beaming}, \tag{56}$$

where $\eta_{\rm BH}$ is the number of BH-forming stars resulting from a given total stellar mass, $\eta_{\rm bin}$ is the binary fraction and $\eta_{\rm close}$ is the fraction of sufficiently close binaries to undergo RLOF. For Pop I/II stars they calculated $\eta_{\rm BH} \simeq 1/(700\ {\rm M}_\odot)$. Combining this value with adopted values for the other parameters—$\eta_{\rm bin} \sim 0.5$, $\eta_{\rm close} \sim 0.3$, and $\eta_{\rm beaming} \simeq (1/50) - (1/500)$—yields $\eta_{\rm GRB,PopI/II} \sim 4.2 \times (10^{-6} - 10^{-7})\ {\rm M}_\odot^{-1}$.

Bromm & Loeb (2006) noted that it was only to make educated guesses for the Pop III case, due to an absence of detailed calculations for the fraction of close binaries. Our work is a first attempt to fill the gap in our theoretical knowledge. We can write $F_{\rm HMXB} = \eta_{\rm BH}\,\eta_{\rm bin}\,\eta_{\rm close}$. Adopting for comparison the same value of $\eta_{\rm beaming} \simeq (1/50) - (1/500)$, we then infer $\eta_{\rm GRB,Pop\ III} \simeq 4.8 \times (10^{-6} \sim 10^{-7})\ {\rm M}_\odot^{-1}$ for the interacting 5-star case, and $2.2 \times (10^{-5} \sim 10^{-6})\ {\rm M}_\odot^{-1}$ for the (most favorable) 10-body scenario.

Therefore, our results suggest that LGRB rates from Pop III stars could be comparable to or somewhat higher than the rates from Pop I/II stars. However, even with these rates, the probability of detecting a GRB from a Pop III star remains very low and rather uncertain. The probability depends on the mass function of Pop III stars, the likelihood that a Pop III star produces a GRB, the luminosity function of these bursts, and how efficiently such high-redshift GRBs can be identified. Accordingly, there is a wide range of published theoretical results based on different model assumptions (Mesinger et al., 2005; Bromm & Loeb, 2006; Salvaterra et al., 2008; Ma et al., 2015). However, the most optimistic estimates suggest that the rate of detection is no higher than one every few decades. Thus, a detection of a GRB from a Pop III star will likely require a long time baseline, provided by a combination of the current *Swift* satellite and proposed future missions such as *SVOM* (Schanne et al., 2010) and *EXIST* (Grindlay, 2009). The latter, in particular, will be especially well-suited for the detection of high-redshift GRBs (see Salvaterra et al. 2008). Observing even a single event from a Pop III star would have profound implications for cosmological models of structure formation (de Souza et al., 2013; Maio & Viel, 2015).

3.4.6 Caveats

Our suite of N-body simulations, spanning diverse sets of initial conditions for Pop III stars and their environments, point that HMXBs form in higher fractions in the earliest galaxies than at low redshifts, and make significant contributions to the thermal history of the IGM, the 21 cm signature at $z \sim 20$, and to the rates of LGRBs. The fact that the formation rates of HMXBs varied little between the simulations suggest that our estimates for $F_{\rm HMXB}$ is reasonably robust. However, we here point out several uncertainties of our work that could affect our conclusions.

One important factor is the IMF of the stars (Hirano et al., 2014). We adopted the IMF of Stacy & Bromm (2013), which were based on the masses of protostars roughly 5000 yr after the formation of the protostellar seeds.

However, our key results are based on fragmentation of the protostellar clouds on smaller scales, as found in the simulations by Greif et al. (2011). In those simulations, the most massive protostars had the highest accretion rates, suggesting that the IMF slope may be

steeper than what we assumed. We also did not account for changes in the masses of stars as they evolved. These are important considerations, as the masses of the stars play a critical role in determining the time that a binary spends transferring mass as a HMXB.

We made an effort to account for this limitation by using free parameters such as $f_{\rm sur}$, which accounts for the mass loss during the SN explosion.

Another factor that could be more carefully treated in future studies is the spin of the star. In a binary system, this impacts the circularization of the orbit, followed by low eccentricity and the synchronization of spin with orbital phase.[9] Due to tidal dissipation and circularization, this could additionally decrease the orbital distance of a binary, boosting the formation of HMXBs and change the average value of the eccentricity.

In addition to reducing the orbital separation, the spin makes a difference in mass transfer rates.

Our simulations generally produce binaries with large eccentricities. For highly eccentric orbits, mass transfer occurs only near pericenter (Lajoie & Sills 2011 and Sepinsky et al. 2010), and the rates depend on whether the orbital angular speed of the star is super-sychronous or sub-synchronous with rotational angular speed (Davis et al., 2013).

We found that the viscous timescales in our eccentric binaries were longer than the orbital timescales in the majority of cases, and used this fact to conclude that their duty cycle should be the same as for nearly circular binaries. However, we did find one exception, in which a RLOF accretion event would be sufficiently short-lived to be episodic. A more careful study focused on stellar rotation effects may be necessary to more conclusively estimate the duty cycle of eccentric Pop III HMXBs.

Finally, our simulations are an attempt to model the stellar dynamics as a gravitational N-body problem with perturbative forces due to a fixed, smooth gaseous background. More detailed simulations that include detailed stellar feedback, as well as the dynamics and thermodynamics of the gaseous environment, could lead to additional revelations about the early evolution of Pop III star groups.

3.5 Summary

In this study, we used N-body simulations of the first stars to explore the formation, evolution, disruption and energy output of Pop III HMXBs. The code includes gravitational scattering of stars, dynamical friction, and the gravitational potential of ambient gas.

The initial conditions for the simulations (i.e. IMF, typical star separation in the host haloes, ambient densities) are taken from two different sets of cosmological simulations of Pop III formation, namely by Stacy & Bromm (2013) ('large-scale', i.e. a few thousands of AU), and Greif et al. (2012), ('small-scale', i.e. a few tens of AU). These provide two complementary sets in that they explore different physical scales for star formation (for details, see 3.2.4). For each of the two scenarios, we investigated star evolution in two background gas densities, a high-density case ($10^6 \mathrm{cm}^{-3}$), and a lower-density one ($10^4 \mathrm{cm}^{-3}$).

Based on the handful of protostars per halo that are found in the works quoted above, we simulated systems with 5 stars and systems with 10 stars. We found:

[9] These effects had been often assumed to be due to the tidal interactions with accreting gas, but recent studies suggest that the orbital semimajor axis and eccentricity can either increase or decrease depending on the binary properties at pericenter (Sepinsky et al. 2007 and Sepinsky et al. 2009).

1. *5-body simulations*: If stars form in quasi-Keplerian disk configurations with initial separations of hundreds of AU, HMXBs are highly unlikely to form. In contrast, if stars form in compact groups separated by $\sim$ 10 AU, as is expected from turbulent fragmentation, stellar scatterings lead to a significant HMXB formation rate. In particular, we found that HMXBs form at a rate of a few per 10^4 $M_\odot$ of stars formed, independent of the ambient gas density.

2. *10-body simulations*: We simulated 10 stars on separations of $\sim$ 10 AU, and evolved them as isolated quasi-Keplerian disks, or as two colliding groups with 5 stars each. For the latter, we ran several different sets of collision geometries.

   We found that the HMXB formation rate was a factor $\sim 1-3$ times higher than for the 5-body simulations, mainly due to the fact that the larger number of stars allowed for more hardening via stellar scattering.

All of the small-scale simulations suggest an X-ray luminosity per unit star formation that is a factor $\sim 10^2$ higher than what is observed in the local Universe (under the assumption that other variables such as the X-ray escape fraction from galaxies and the duty cycle of HMXBs do not differ significantly). These results are mostly due to *the large mass of the most massive star of the HMXBc compared to that of the companion star, implying both a large* $t_{\rm acc}$ *as well as a higher luminosity of the remnant BH*. The fact that we found little variation in this quantity across all of our simulations suggests that this is a robust estimate. Additional factors, such as in-disk migration of nascent stars, could further increase the HMXB formation efficiency.

A direct consequence is that X-rays can heat the IGM rapidly at Cosmic Dawn. Signals of early heating can be probed via the 21 cm line radiation: the absorption line signal is expected to show a broader, shallower minimum due to the shorter gas cooling time, while the emission line would be observed earlier because of the higher gas temperature, at earlier times. Several studies have modeled the 21 cm signature of the first HMXBs, but relied on assumptions for their $L_{\rm X}$/SFR relation relative to the empirical value found at lower redshifts. Our work provides a theoretically driven estimate for this quantity.

In addition to the implications for the thermal history of the IGM, these high formation rates of HMXBs per stellar mass imply a higher GRB formation efficiency from Pop III stars in binaries. These predictions may be tested with a long baseline of observational data from *Swift* in combination with future missions such as *SVOM, EXIST*.

4 Intermediate-mass black holes from Population III remnants in the first galactic nuclei

T. Ryu, T. Tanaka, R. Perna, Z. Haiman *The Monthly Notices of the Royal Astronomical Society*, Vol. 460, 4122 (2016)

Abstract

We report the formation of intermediate-mass black holes (IMBHs) in suites of numerical N-body simulations of Population III remnant black holes (BHs) embedded in gas-rich protogalaxies at redshifts $z \gtrsim 10$. We model the effects of gas drag on the BHs' orbits, and allow BHs to grow via gas accretion, including a mode of hyper-Eddington accretion in which photon trapping and rapid gas inflow suppress any negative radiative feedback. Most initial BH configurations lead to the formation of one (but never more than one) IMBH in the center of the protogalaxy, reaching a mass of 10^{3-5} $M_\odot$ through hyper-Eddington growth. Our results suggest a viable pathway to forming the earliest massive BHs in the centers of early galaxies. We also find that the nuclear IMBH typically captures a stellar-mass BH companion, making these systems observable in gravitational waves as extreme mass-ratio inspirals (EMRIs) with *eLISA*.

4.1 Introduction

Virtually every nearby massive galaxy harbors a supermassive black hole (SMBH) in its nucleus (Kormendy & Ho, 2013; Magorrian et al., 1996). Tight correlations between SMBHs and the properties of their host galaxies, as well as the phenomenology of quasars and luminous active galactic nuclei, suggest that SMBHs play a role in shaping their environment both on local (Silk & Rees, 1998; Fabian, 2012) and cosmological (Madau et al., 2004; Ricotti & Ostriker, 2004b; Tanaka et al., 2016) scales.

How SMBHs form is one of the most fundamental open problems in astrophysics. The observation of quasars powered by $\sim 10^9$ $M_\odot$ SMBHs at $z \approx 6-7$, less than a Gyr after the Big Bang (Fan et al., 2001; Willott et al., 2003, 2010; Mortlock et al., 2011; Venemans et al., 2013), places strong constraints on theoretical models for their origin (see reviews by, e.g. Volonteri 2010; Haiman 2013). The two most often discussed hypotheses are that they grew from either (i) the stellar BH remnants of the first generation of ~ 100 $M_\odot$ stars (Population III, hereafter Pop III, stars; Haiman & Loeb 2001; Madau & Rees 2001; Yoo & Miralda-Escudé 2004; Volonteri & Rees 2006; Tanaka & Haiman 2009), or (ii) the remnants of the "direct collapse" of $> 10^5$ $M_\odot$ gas clouds that avoided fragmentation into Pop III stars (Oh & Haiman, 2002; Bromm & Loeb, 2003; Koushiappas et al., 2004; Volonteri & Rees, 2005; Shapiro, 2005; Begelman et al., 2006; Spaans & Silk, 2006; Lodato & Natarajan, 2006; Wise & Abel, 2008; Regan & Haehnelt, 2009; Schleicher et al., 2010; Shang et al., 2010; Latif et al., 2013; Tanaka & Li, 2014). Both scenarios require that some of the first SMBHs grew at a (logarithmically averaged) rate $\dot{M} \sim 10 L_{\rm Edd}/c^2$ (Tanaka, 2014), where $L_{\rm Edd} \propto M$ is the Eddington luminosity for a BH of mass M, and c is the speed of light. This value is comparable to the accretion rate producing the Eddington luminosity with the radiative

efficiency $\eta \equiv L/(\dot{M}c^2) \sim 0.1$ expected in thin discs.[10]

In this paper, we examine the possibility that the first nuclear SMBHs originated from *hyper-Eddington* accretion onto Pop III remnants—i.e. from a growth mode where $\dot{M} \gg \dot{M}_{\rm Edd} \equiv L_{\rm Edd}/c^2$. This is motivated by theoretical models of optically thick accretion flows in which photons are trapped and advected inside the accretion flow. In such "radiatively inefficient" accretion modes the luminosity and radiation feedback of the flow are quenched, allowing accretion rates much higher than those corresponding to the Eddington limit for radiatively efficient discs (see Begelman 1979 for spherical flows and Abramowicz et al. 1988 for slim accretion discs). Several studies have investigated whether such an accretion mode contributed to the growth of the first SMBHs (Volonteri & Rees, 2005; Begelman, 2012; Madau et al., 2014; Alexander & Natarajan, 2014; Volonteri et al., 2015; Pacucci & Ferrara, 2015; Pacucci, Volonteri, & Ferrara, 2015; Lupi et al., 2016; Pezzulli et al., 2016).

We focus our attention on the recent work by Inayoshi et al. (2016), who found hyper-Eddington accretion solutions in spherically symmetric radiation-hydrodynamics simulations of Bondi-like accretion. They found, in broad agreement with previous simulations (Milosavljević et al., 2009; Park & Ricotti, 2012), that radiative feedback typically limits the accretion rate to values comparable or below $\dot{M}_{\rm Edd}$. This radiative feedback arises from photoionization and heating of the gas near the Bondi radius, and occurs even for flows that are highly optically thick to electron scattering, and for which trapping of the radiation limits the luminosity emerging from the photosphere below $L_{\rm Edd}$. However, Inayoshi et al. (2016) also found that for sufficiently large ambient gas densities, the combination of the large ram pressure of the inflowing gas and photon trapping inhibit radiative feedback. The accretion flow in this regime is steady and unimpeded from the Bondi rate; following Inayoshi et al. (2016) we refer to this as hyper-Eddington accretion.

We perform N-body simulations of Pop III remnant BHs in a model protogalactic distribution of gas and dark matter (DM) at $z \gtrsim 10$, subjecting the BHs to dynamical friction, and allowing them to accrete in the manner found by Inayoshi et al. (2016). The goal of our simulations is to follow the coupled growth and orbital evolution of a small cluster of stellar-remnant BHs, and to evaluate whether such BHs can reasonably be expected to reach the hyper-Eddington regime, and grow rapidly into more massive BHs.

In our models, we find that Pop III BHs indeed frequently grow into IMBHs with masses over 10^3 $M_\odot$, and even into supermassive holes (SMBHs) with masses 10^5 $M_\odot$. We further find that these I/SMBHs always form after a lower-mass BH has eroded its orbit, and settled near the center of the model protogalaxy. This suggests that hyper-Eddington accretion is a viable mechanism for forming nuclear SMBHs in early galaxies.

In addition, we report that our simulations always produce only a single I/SMBH. This is because once these massive BHs dominate the central potential, subsequent BHs dragged into the dense central regions reach high velocities that prohibit their growth. Instead, they are captured as stellar-mass BHs in a bound orbit. This suggests that I/SMBH formation is typically accompanied by so-called extreme-mass ratio inspirals (EMRIs), the merger of compact objects with mass ratios $\ll 1$ that are one of the main targets of the planned

[10]Radiatively efficient discs are theoretically expected to have $\eta \sim 0.1 - 0.4$ (e.g. Shapiro, 2005), in agreement with the mean value $\sim 0.07 - 0.1$ inferred from the Sołtan-Paczynski argument, comparing the mass density of SMBHs with the quasar luminosity density (e.g. Merloni & Heinz 2008; Shankar et al. 2010, but see Shankar et al. 2016 for a recent argument for a higher value).

space-borne gravitational-wave detector *eLISA*.

The rest of this paper is organized as follows. In §2, we describe the setup of our simulations, including the properties of the Pop III remnant BHs and of their host galaxy, as well as the numerical schemes used to simulate their orbital evolution and growth. We present and discuss our main results in §3. Several implications and theoretical caveats are discussed further in §4. We summarize our conclusions in §5.

4.2 Numerical model

In this section, we provide an overview of our simulations. We first describe the properties of the parent galaxy, and the initial conditions for the small cluster of stellar-mass BHs. We then describe the equations of motion we solve to follow the growth of BHs and their interactions and dynamics, as well as the numerical scheme we use to solve these equations. A key feature of our model is a prescription that allows rapid growth by gas accretion, based on the recent numerical study of Bondi-like hyper-Eddington accretion with radiation by Inayoshi et al. (2016).

4.3 Protogalaxy + BH population model

4.3.1 Protogalactic gas cloud and DM halo

We consider a small cluster of Pop III-remnant BHs embedded in a protogalactic, so-called atomic-cooling DM halo, with a virial temperature of $T_{\rm vir} \geq 10^4$K and mass $\gtrsim 10^{7-8}$ $M_\odot$ at $z \sim 15-20$. Such a system is a plausible outcome of lower-mass Pop III-forming haloes growing either via accretion and minor mergers, or via major mergers. Massive stars are expected to form earlier, in the lower-mass progenitor "minihaloes", and leave behind stellar-mass BH remnants. However, the UV radiation and/or SN explosions of the progenitor star of such a BH can unbind the gas from the shallow potential well of its host minihalo. The remnant BHs are then expected to be starved Alvarez et al. (2009), and not surrounded again by dense gas and grow until they are incorporated into more massive atomic-cooling haloes.

The gas collapsing inside an atomic cooling halo is expected to cool and condense, and to develop a nearly isothermal density profile with $\mathrm{d}\ln\rho/\mathrm{d}\ln r \approx -2$ (Oh & Haiman, 2002; Wise & Abel, 2007; Regan & Haehnelt, 2009; Shang et al., 2010; Latif et al., 2014; Regan et al., 2014). Under the assumption that the metallicity in this protogalaxy remains low, and H_2 cooling is disabled, the temperature will remain near the HI atomic cooling floor of 8000 K. This configuration is expected to be rare, as it requires a large UV (Lyman-Werner) flux from a bright neighbour, forming near-simultaneously (Dijkstra et al., 2008; Agarwal et al., 2012; Visbal et al., 2014). In the presence of metal and/or H_2-cooling, the large-scale inflow rate is expected to be slower, as a result of the lower sound speed, reducing the normalization of the density profile (see, e.g. Shang et al. 2010).

Our fiducial protogalaxy model is based on the metal- and H_2–free atomic cooling halo, and consists of a DM component with a Navarro et al. (1997) (NFW) density profile, and an isothermal gas profile that behaves as $\propto r^{-2}$ at large radii. In order to avoid a mathematical

singularity at the origin, we introduce a core region of size r_c inside which the density is nearly constant.

The matter distribution in our model protogalaxy can be summarized as follows:

$$\rho_{bg}(r) = \rho_{gas}(r) + \rho_{NFW}(r) \tag{57}$$

$$\rho_{gas}(r) = \frac{\rho_c}{1 + (\frac{r}{r_c})^2}, \tag{58}$$

where $\rho_{NFW}(r)$ is the NFW profile with concentration parameter $C = 9$ and the virial radius is $r_{vir} \simeq 1$ kpc. The virial radius is defined as the radius within which the average matter density is 180 times the cosmological critical density. The values of r_c and n_c are determined by normalizing the gas profile to satisfy the cosmological ratio of baryon to DM mass inside r_{vir}. We take $r_c = 0.003$ pc and $n_c = 2.5 \times 10^{10}\text{cm}^{-3}$. These values are consistent with those found in the highest-resolution adaptive mesh refinement simulations to date (Regan et al., 2014). For numerical convenience, we slightly modify the NFW profile (which scales as r^{-1} at small radii) by requiring that the DM density does not exceed the gas density for $r \leq r_{NFW,c}$, a radius at which $\rho_{gas} = \rho_{NFW}$. This modification only affects the region inside $\sim 10^{-3} r_c$, and does not appreciably affect our results. The density profile of DM and gas remains fixed and unchanging throughout our simulations; the possible impact of this large simplification is addressed below.

4.3.2 Pop III-remnant BHs

Within our spherically symmetric halo, we place a small cluster of ten BHs inside a radius of 100 pc. This represents $\approx 10\%$ of the virial radius of a halo just above the atomic cooling threshold, and is intended to correspond to the spatial extent of a star-forming region, or the region over which the BHs are initially spread after merging events. Each hemisphere is divided by the polar angle into five compartments of equal shape and size, and a single BH is placed at a randomly chosen radius and angular position inside each compartment (i.e., ten compartments, and one BH per compartment).

We also require that the initial distance between each pair of BHs is larger than 10 pc. Each BH is given a random initial velocity, so that its speed is no larger than the gas sound speed $c_s \approx 10\,\text{km}\,\text{s}^{-1}$ and the radial component of its velocity is nonpositive (i.e. it is not flying away from the origin; this is intended to mimic the outcome of recent mergers).

The masses of the Pop III stars as the BH progenitors are assigned randomly from an IMF $\frac{dN}{dM_\star} = M_\star^{-\alpha}$ with $\alpha = 0.17$ Stacy & Bromm 2013. We adopt a minimum mass of $M_{\text{min},\star} = 25$ M$_\odot$, and a maximum mass of $M_{\text{max},\star} = 140$ M$_\odot$. The latter value is motivated by the fact that more massive stars are expected to result in pair-instability supernovae and leave no BH remnant Heger et al. 2003. While stars more massive than ≈ 260 M$_\odot$ may form BHs, such large masses may be precluded by UV feedback in the protostellar stages Hosokawa et al. e.g. 2011.

After the stellar masses are drawn, they are converted to BH masses M using the following fitting formulae provided by Tanaka et al. (2012b), which are based on simulations by Zhang

Redshift	$z \approx 15 - 20$
Sound speed	$c_{\rm s} \approx 10$ km s^{-1}
Halo mass	$M_{\rm vir,halo} \approx 5 \times 10^7 - 10^8$ M$_\odot$
Halo virial temperature	$T_{\rm vir} \gtrsim 10000$ K
Initial radial distance of BHs	$r_i(t = 0) \leq 100$ pc
Initial separations between BHs	$r_{ij}(t = 0) \geq 10$ pc
Initial speed of BHs	$v_i(t = 0) \leq c_{\rm s}$
Mean molecular weight	$\mu_{\rm m} = 1$
Gas density profile	$\rho_{\rm gas} \propto r^{-2}$ (isothermal sphere)
Dark matter density profile	$\rho_{\rm DM}$=NFW
Concentration parameter	C=9
Virial radius	$r_{\rm vir} \simeq 1$ kpc
Core gas density	$n_{\rm c} = 2.5 \times 10^{10}$ cm^{-3}
Core radius for gas	$r_{\rm c} = 3 \times 10^{-3}$ pc
Core radius for DM	$r_{\rm NFW,c} \approx 10^{-6}$ pc
Initial stellar mass function range	25 M$_\odot \leq M_\star \leq$ 140 M$_\odot$
Initial stellar mass function slope	$\alpha = 0.17$
Total run time	$t_{\rm run} = 500$ Myrs

Table 8: Choices of the values of different physical parameters defining our protogalaxy + BH cluster model. See text for details. Note that the italic subscripts i and j are indices representing the BHs.

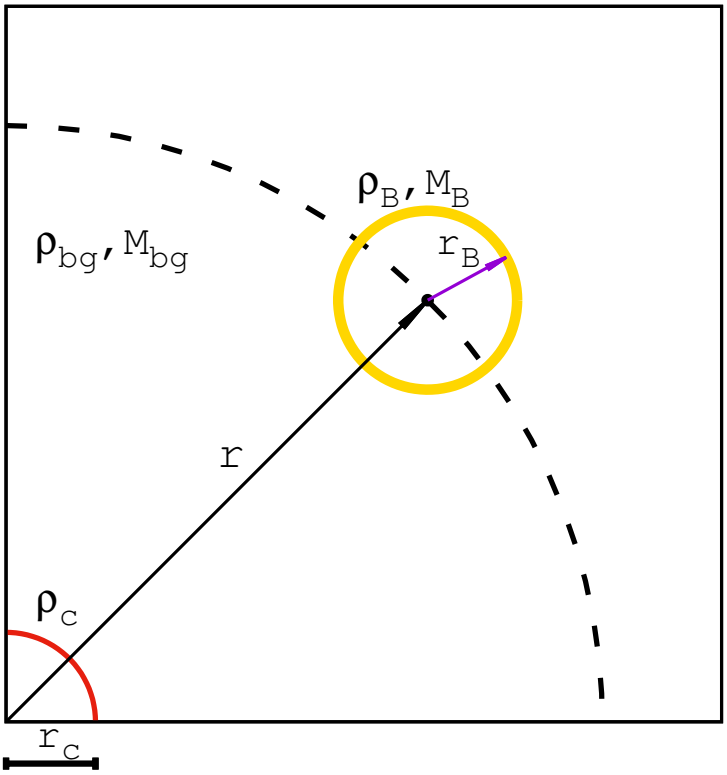

Figure 19: Schematic representation of one of the BHs in the halo, defining our notation. $M_{\rm bg}(< r)$ denotes the mass of the background material (gas + DM) enclosed within the radial position r of the BH, while $M_{\rm B}$ represents the mass of gas enclosed within the Bondi radius of the BH.

et al. (2008):

$$M = \begin{cases} \frac{3}{4}(M_\star - 20\ {\rm M}_\odot) + 2\ {\rm M}_\odot & \text{if } M_\star \leq 45\ {\rm M}_\odot \\ \frac{5}{12}(M_\star - 20\ {\rm M}_\odot) & \text{if } M_\star > 45\ {\rm M}_\odot\,. \end{cases} \tag{59}$$

The average mass of a Pop III star is $\langle M_\star \rangle \approx 80\ {\rm M}_\odot$, and that of a remnant BH is $\langle M \rangle \approx 25\ {\rm M}_\odot$. All simulations are run until either a tightly bound BH pair with a semimajor axis $a \leq 1$ pc forms, or until a physical time of $t_{\rm run} = 500$ Myr has elapsed—whichever occurs first.

The properties of the host halo and the BHs adopted in our simulations are summarized in Table 8, and a schematic diagram of the halo + BH system is illustrated in Figure 19. The radial coordinate of the BH is denoted by r. The gas density profile is spherically symmetric, and has a flat central core of size $r_{\rm c}$. The sum of the gas and DM densities define the total background matter density $\rho_{\rm bg}$. We denote the mass enclosed inside the instantaneous radial coordinate of the BH by $M_{\rm bg}(< r)$. The Bondi radius $r_{\rm B}$ (defined in §4.5 below) defines a spherical region around each BH that we call the "Bondi sphere." We will make use of the average density around the surface of this sphere, $\langle \rho \rangle_{\rm B}$, and the total mass enclosed inside the Bondi sphere, $M_{\rm B}$.

4.4 The equations of motion

We use an N-body code to integrate the equations of motion and mass growth for each BH embedded in the model protogalaxy. The motion of the BHs is determined by the following forces: (i) their mutual gravitational attraction (including post-Newtonian terms up to 2.5th order), (ii) dynamical friction/drag from the surrounding medium, and (iii) the gravitational potential of the background matter. We also account for (iv) the deceleration due to mass growth via accretion. The resulting equation of motion for the $i^{\rm th}$ BH includes the sum of the five forces:

$$\mathbf{a}_i = \mathbf{a}_{{\rm N},i} + \mathbf{a}_{{\rm PN},i} + \mathbf{a}_{{\rm df},i} + \mathbf{a}_{{\rm bg},i} + \mathbf{a}_{{\rm acc},i} \tag{60}$$

We next describe each contribution in detail.

1. *Gravitational attraction between BHs*
   This contribution is dominated by the standard Newtonian expression,

$$\mathbf{a}_{{\rm N},i} = -\sum_{j\neq i} G\ M_{{\rm BH},j}\ \frac{\partial\ \Phi(r_{ij})}{\partial\ r_{ij}}\ \frac{\mathbf{r}_i - \mathbf{r}_j}{r_{ij}}, \tag{61}$$

   where G is the gravitational constant, Φ is the pairwise gravitational potential, $\mathbf{r}_i$ is the displacement of the $i^{\rm th}$ BH from the center of the host DM halo, and $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$. In our numerical implementation, we adopt the Plummer softening kernel Binney & Tremaine 1987 with softening length 1.7×10^6 cm, which is equivalent to the Schwarzschild radius for a 5.75 $\rm M_\odot$ BH.

   We add post-Newtonian terms $\mathbf{a}_{\rm PN}$ to Eq. (61) up to order 2.5, which includes the loss of orbital energy and angular momentum via GWs. The full expressions for these terms can be found in, e.g., (Damour & Deruelle, 1981; Kupi et al., 2006). At sufficiently small pair separations, the orbital decay due to GW emission leads to merger on a timescale $\propto a^4$ where a is the semimajor axis (Peters & Mathews, 1963; Peters, 1964). It turns out that no BHs in our simulations reach separations where GW emission is relevant for their orbital evolution[11] Notice that This finding is different from the similar, earlier studies by Tagawa et al. (2015, 2016), who considered a cluster of BHs with a smaller initial separation, and found that the post-Newtonian terms can sometimes be important and lead to BH-BH mergers. We defer the discussion of this point to § 4.8 below.

2. *Dynamical friction and gas drag*
   An object in motion through a medium creates an overdensity, or wake, behind it, whose gravitational pull acts as a dissipative drag on the object's motion. In this study, we consider dynamical friction due to both DM and gas.

[11] For reference, GW emission becomes the dominant orbital evolution mechanism at an orbital distance $a < 10^{-5}$ pc for a circular binary consisting of two BHs with masses 10^5 $\rm M_\odot$ and 100 $\rm M_\odot$. We find that the central pair is rarely disrupted once they reach a separation ~ 1 pc, as described later in the text (§4.7.3).

For the DM contribution, we adopt the standard Chandrasekhar formula (Binney & Tremaine, 1987),

$$\mathbf{a}_{\mathrm{df},i}^{(\mathrm{DM})} = -4\pi \ \ln\Lambda \ f(X_i) \ \frac{G^2 M_i}{v_i^3} \ \rho_{\mathrm{DM}}(r_i) \ \mathbf{v}_i, \tag{62}$$

with

$$f(X_i) \equiv \mathrm{erf}(X_i) - \frac{2}{\sqrt{\pi}} \ X_i \ \exp\left(-X_i^2\right), \tag{63}$$

where $X_i \equiv v_i/(\sqrt{2}\sigma_v)$ and σ is the velocity dispersion, $\simeq c_{\mathrm{s}}$. We use $\ln\Lambda = 3.1$ and indicate with $\rho_{\mathrm{DM}}(r_i)$ the DM density at the location of the i-th BH.

For the gas, we adopt the following formula from Tanaka & Haiman (2009), which incorporates behaviours found in numerical simulations for subsonic and supersonic regimes (Ostriker, 1999b; Escala et al., 2004; Chapon et al., 2013). In our implementation, the specific drag force vector always points opposite to the direction of BH motion, and is given by:

$$\mathbf{a}_{\mathrm{df},i}^{(\mathrm{gas})} = -4\pi \ G^2 \ M_i \ \rho_{\mathrm{gas}}(r_i) \ \frac{1}{v_i^3} \times f^{(\mathrm{gas})}(\mathcal{M}_i)\mathbf{v}_i, \tag{64}$$

with

$$f^{(\mathrm{gas})}(\mathcal{M}_i) = \begin{cases} 0.5 \ \ln\Lambda\left[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i \ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\right] \\ \qquad\qquad 0 \leqslant \mathcal{M}_i \leqslant 0.8; \\ 1.5 \ \ln\Lambda\left[\mathrm{erf}\left(\frac{\mathcal{M}_i}{\sqrt{2}}\right) - \sqrt{\frac{2}{\pi}}\mathcal{M}_i \ \exp\left(-\frac{\mathcal{M}_i^2}{2}\right)\right] \\ \qquad\qquad 0.8 \leqslant \mathcal{M} \leqslant \mathcal{M}_{\mathrm{eq}}; \\ \frac{1}{2}\ln\left(1 - \frac{1}{\mathcal{M}_i^2}\right) + \ln\Lambda \\ \qquad\qquad\qquad\qquad \mathcal{M}_i > \mathcal{M}_{\mathrm{eq}}. \end{cases} \tag{65}$$

Above, $\mathcal{M}_i \equiv v_i/c_{\mathrm{s}}$ is the Mach number of the ith BH, and c_{s} is the isothermal sound speed of the gas. We use $\ln\Lambda = 3.1$, the same as for the DM. The corresponding value of $\mathcal{M}_{\mathrm{eq}}$ that makes the above function continuous with respect to $\mathcal{M}$ is approximately 1.8.

With the density distributions we use in our model (see § 4.3 above), and in particular near the center of the model protogalaxy, the effects of gas dominate over that of DM—both in dynamical friction and background gravitational force. Although we include the DM-related force calculations for completeness, they do not play a major dynamical role.

The expressions for dynamical friction given above were derived under the assumption of non-accelerated motion in a uniform density distribution. Capturing the effects of nonlinear dynamical friction along an accelerated trajectory, in a non-uniform background medium, would require hydrodynamical simulations, including the self-gravity of the surrounding medium. This is outside the scope of this paper, but we note that existing studies of dynamical friction in a nonuniform medium or for perturbers on nonlinear trajectories (e.g Sánchez-Salcedo & Brandenburg, 2001; Just & Peñarrubia,

2005; H. Kim & Kim, 2007; W.-T. Kim, 2010), do not report major differences from the Chandrasekhar formula. We therefore simply evaluate the formulae given above, by using the value of the density at the coordinates of each BH; this is the typical approach taken in similar numerical studies (e.g. Blecha et al., 2011; Guedes et al., 2011). For comparison, we have run a second set of simulations in which the dynamical friction forces were computed by averaging the density values at a distance around each BH; we defer discussing the details of this comparison until §4.6 below.

3. *Gravity of the background matter*
The background gas and DM exert a gravitational pull on the BHs. Because we assume a spherically symmetric density profile, this force points toward the center of the potential. It can be expressed as

$$\mathbf{a}_{\mathrm{bg},i} = -\frac{G\ M_{\mathrm{bg},i}}{r_i^3}\mathbf{r}_i, \tag{66}$$

where $\mathbf{r}_i$ is vector pointing from the center of the halo to the i-th BH and $M_{\mathrm{bg},i}$ is the mass of ambient gas and DM contained inside $r < r_i$.

Our assumption that the background matter distribution remains static will fail when $M \gtrsim M_{\mathrm{bg}}$, i.e. when the BH mass exceeds that of the matter inside its orbit. In this case, the matter at the center will be strongly perturbed by the BH, and our prescription of a static background is invalidated. Treating the dynamical reaction of the gas and DM distribution to a massive BH is beyond the scope of our computational methods. However, in order to assess the possible impact of this assumption on our results, we have run two sets of simulations with different treatments of the background force. In the first set, we treat this force simply as given by equation 66. In the second, we set it to zero if $M > M_{\mathrm{bg}}$. The full set of our simulations is described in §4.6.

4. *Accretion-induced deceleration*
The BH decelerates through conservation of linear momentum,

$$\mathbf{a}_{\mathrm{acc},i} = -\frac{\dot{M}_i}{M_i}\mathbf{v}_i. \tag{67}$$

4.5 Accretion rate

We next detail our prescription for the mass growth of each BH due to gas accretion. We base our model on the recent numerical study by Inayoshi et al. (2016), whose key finding is that the accretion rate can significantly exceed the Eddington rate (see also the other references mentioned in the Introduction). Inayoshi et al. (2016) found that spherically symmetric BH accretion solutions with radiative feedback were divided into several qualitatively distinct

regimes, depending on the ratio of the Bondi accretion rate of ambient gas (ρ_∞),

$$\dot{M}_{\rm B} = \frac{4\pi G^2 M^2 \ \rho_\infty}{c_{\rm s,\infty}^3}, \tag{68}$$

($c_{\rm s,\infty}$ being the sound of speed of the ambient gas), to the Eddington rate[12]

$$\dot{M}_{\rm Edd} \equiv \frac{L_{\rm Edd}}{c^2} = \frac{4\pi G M}{\kappa_{\rm es}\ c} = 2.2 \times 10^{-9} \frac{M}{\rm M_\odot}\ {\rm M_\odot}\ \ {\rm yr}^{-1}, \tag{69}$$

where $L_{\rm Edd}$ is the Eddington luminosity and $\kappa_{\rm es}$ is the Thomson scattering opacity.

Inayoshi et al. (2016) concluded that:
(i) Under conditions where the ratio of the canonical Bondi rate to the Eddington rate $\dot{M}_{\rm B}/\dot{M}_{\rm Edd} \lesssim 0.1 - 1$, the BH accretion rate is $\dot{M} \sim \dot{M}_{\rm B}$;
(ii) If $1 \lesssim \dot{M}_{\rm B}/\dot{M}_{\rm Edd} \lesssim 100$, photoionization by the light produced by the accretion flow causes the accretion onto the BH to be intermittent, with a time-averaged rate $\dot{M} \lesssim \dot{M}_{\rm Edd}$;
(iii) If $\dot{M}_{\rm B}/\dot{M}_{\rm Edd} \gtrsim 3000$, the large ram pressure of the inflowing gas, combined with photon trapping below the photosphere, renders radiative feedback ineffective, and accretion proceeds unimpeded at $\dot{M} \sim \dot{M}_{\rm B}$ (cf. Begelman, 2012).
In the intermediate regime between cases (ii) and (iii) above, i.e. if $100 \lesssim \dot{M}_{\rm B}/\dot{M}_{\rm Edd} \lesssim 3000$, the accretion rate remains uncertain, because of the unresolved role of hydrodynamical instabilities.

Our accretion prescription closely follows the behaviour outlined above, with a few modifications. First, we replace the Bondi rate for a stationary mass with the Bondi-Hoyle-Lyttleton (BHL) rate, to account for the fact that the BHs in our simulation are in motion with respect to the surrounding gas. Second, whereas the canonical expression for the Bondi accretion rate (eq. 68) uses the ambient density "at infinity" ρ_∞, we instead use the value of the density averaged over the spherical region around the BH defined by the Bondi radius

$$r_{{\rm B},i} = \frac{2GM_i}{c_{\rm s}^2 + v_i^2}. \tag{70}$$

The resulting expression for our modified Bondi rate is

$$\dot{M}_{{\rm B},i} = \frac{4\pi G^2 M_i^2 \ \langle \rho_{\rm B}(r_i) \rangle}{c_{\rm s}^3 (1 + \mathcal{M}_i^2)^{3/2}}, \tag{71}$$

where $\langle \rho_{\rm B}(r_i) \rangle$ denotes the aforementioned average density of gas at the surface of the "Bondi sphere" around the BH.

Third, we conservatively assume that the BH accretion rate should not be higher than the mass inflow rate into the center of the halo from larger scales, as this would deplete the central gas density, without the possibility of a steady replenishment from larger radii. In pristine atomic–cooling haloes, the hydrodynamical simulations mentioned above typically

[12]Our notation was chosen to match that of Inayoshi et al. (2016). Many other works define $\dot{M}_{\rm Edd}$ to correspond to the Eddington luminosity with a radiative efficiency of η, which would be $\dot{M}_{\rm Edd}/\eta$, i.e. a factor of 10 higher than eq. (69) for $\eta = 0.1$.

Table 9: Summary of our 6 prescriptions for BH accretion and 2 different treatments of the central background potential adopted in our simulations. These constitute a set of 12 models. Note that the accretion prescription in Eq. 73 follows Inayoshi et al. (2016). In model "I" this prescription is adopted independently of whether the BH mass exceeds or not the mass contained within its Bondi radius, $M_{\rm B}$, whereas in the "$f_{\rm in}$" models the accretion rate is capped to a fraction $f_{\rm in}$ of the inflow rate in the self-gravitating regime, when $M_{\rm B} \geq M$. Model "E" denotes a commonly adopted Eddington-limited accretion prescription, and in the $\dot{M} = 0$ reference case we do not allow any accretion.

Model		$\dot{M}$ when $M_{\rm B} < M$	$\dot{M}$ when $M_{\rm B} \geq M$
Accretion Prescription	$f_{\rm in} = 1$	eq.73	$\dot{M} = f_{\rm in}\dot{M}_{\rm in} = \dot{M}_{\rm in}$
	$f_{\rm in} = 10^{-3}$	eq.73	$\dot{M} = 10^{-3} \times \dot{M}_{\rm in}$
	$f_{\rm in} = 0$	eq.73	$\dot{M} = 0$
	I	eq.73	eq.73
	E	$\min[\dot{M}_{\rm B}, \frac{1}{\eta}\dot{M}_{\rm Edd}, \dot{M}_{\rm in}]$	$\min[\dot{M}_{\rm B}, \frac{1}{\eta}\dot{M}_{\rm Edd}, \dot{M}_{\rm in}]$
	$\dot{M} = 0$	$\dot{M} = 0$	$\dot{M} = 0$
Background Potential	On	$a_{{\rm bg},i}$ always on	
	Off	$a_{{\rm bg},i} = 0$ if $M_i > M_{\rm bg}(< r_i)$	

find this inflow rate to be

$$\dot{M}_{\rm in} = \frac{c_{\rm s}^3}{G} \approx 0.5 \ {\rm M}_\odot \ {\rm yr}^{-1}. \tag{72}$$

A possible caveat here is that the inflow rate in the presence of metal and/or H_2 cooling may be $\approx$ two orders of magnitude lower. On the other hand, the rate increases steadily as the halo grows in mass, and recent simulations have found that pressure and gravitational torques can maintain $\sim$ $\rm M_\odot$ $\rm yr^{-1}$ inflow rates down to $\sim$pc scales, even in the face of radiative cooling and SN feedback (Prieto & Escala, 2016).

Our implementation of the Inayoshi et al. (2016) accretion regimes can therefore be summarized as:

$$\dot{M} = \begin{cases} \min[\dot{M}_{\rm B}, \frac{1}{\eta}\dot{M}_{\rm Edd}, \dot{M}_{\rm in}] \\ \qquad\qquad (\text{if } \min[\dot{M}_{\rm B}, \dot{M}_{\rm in}] < 3000\dot{M}_{\rm Edd}) \\ \\ \min[\dot{M}_{\rm B}, \dot{M}_{\rm in}] \\ \qquad\qquad (\text{if } \min[\dot{M}_{\rm B}, \dot{M}_{\rm in}] \geq 3000\dot{M}_{\rm Edd}) \end{cases} . \tag{73}$$

In this study, we take $\eta = 0.1$. We note that this may somewhat overestimate the accretion rate in the Eddington-limited regime, as both Park & Ricotti (2012) and Inayoshi et al. (2016) found that the time-averaged rate is limited to $\sim 0.5\dot{M}_{\rm Edd}$. We implement

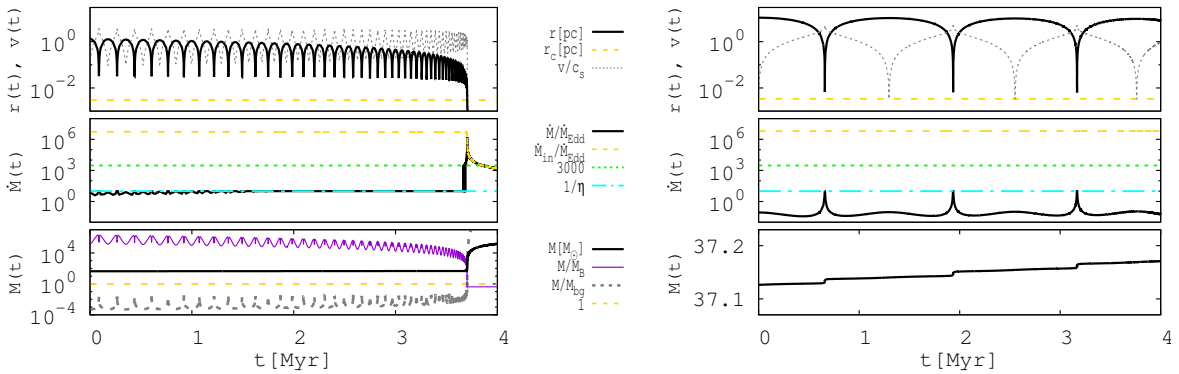

Figure 20: *Top left panel:* The radial distance r (black solid line), and Mach number (gray dotted line) of a BH which experiences the hyper-Eddington accretion as it sinks to the halo core, whose size r_c is also displayed for reference (yellow dashed line) in Model "$f_{\rm in} = 1$" with $a_{\rm bg}$ 'On' (see Table 9). *Middle left panel:* The corresponding accretion rate $\dot{M}$ of the BH in Eddington units (solid black line). For reference, the panel also shows the the self-gravitating inflow rate $\dot{M}_{\rm in}$ (yellow dashed) also in Eddington units, as well as the critical values at which the accretion switches modes [i.e. to Eddington (blue dot-dashed) or hyper-Eddington (green dotted)]. *Bottom left panel:* The corresponding mass of the BH is shown (solid black line), together with the mass $M_{\rm B}$ enclosed within its Bondi radius (solid magenta) and the gas mass $M_{\rm bg}$ contained inside its orbit (dashed gray). The accretion rate and the BH mass both grow rapidly once the BH sinks the core. At the same time, the BH begins to dominate the central potential, but not the mass inside its Bondi sphere. *Right panels:* The above behaviour is contrasted with a BH on a highly elliptical orbit but with a larger semimajor axis, which never makes it inside the core. This BH never experiences (hyper-)Eddington accretion, and its mass remains near its initial value.

Equation (73) as long as the BH dominates the gravitational potential inside its Bondi sphere, i.e. $M > M_{\rm B}$. On the other hand, if $M < M_{\rm B}$, the Bondi formalism breaks down. The latter condition roughly coincides with the gas inside the Bondi sphere becoming self-gravitating and Jeans-unstable. In this regime, the accretion rate is plausibly of order $\dot{M}_{\rm in} \approx c_{\rm s}^3/G$. However, how much of this canonical inflow rate ends up accreting onto the BH remains uncertain, and will depend on factors such as gas cooling, turbulence and angular momentum transport, and the BH's specific accelerated trajectory. We here parameterize the accretion rate in this self-gravitating gas regime as $\dot{M} = f_{\rm in}\dot{M}_{\rm in}$, and consider the two extreme values of $f_{\rm in} = 0$ and 1, as well as an intermediate value of $f_{\rm in} = 10^{-3}$.

To further explore the dependence of our results on this accretion prescription, we have also run a set of simulations where $\dot{M}$ continues to be given by the Bondi rate even when $M_{\rm B} > M$, and another set where $\dot{M}$ continues to follow the Inayoshi et al. (2016) prescription described above, regardless of whether $M_{\rm B}$ is larger or smaller than M.

4.6 Summary of simulation sets

Our fiducial set of models implement the prescriptions described above. The accretion rate onto the BH is determined by the Inayoshi et al. (2016) accretion rates (eq. 73) if $M > M_{\rm B}$,

and $\dot{M} = f_{\rm in} M_{\rm in}$ in the self-gravitating regime, once $M \leq M_{\rm B}$. As noted above, we have run two additional sets of simulations, in which $\dot{M}$ is either given by Eq. (73) regardless of how M compares to $M_{\rm B}$ (which we will refer to as (Prescription "I"; for Inayoshi et al.) and another in which BH accretion tracks the Bondi rate, but is always Eddington-limited, i.e. $\dot{M} = \min[\dot{M}_{\rm B}, \frac{1}{\eta}\dot{M}_{\rm Edd}, \dot{M}_{\rm in}]$ (Prescription "E", for Eddington). Note that the latter is a prescription commonly adopted in numerical and semi-analytic studies of BH growth in the early Universe (see reviews by, e.g. Volonteri 2010; Haiman 2013 and references therein) As a simple control, we have also run simulations with no accretion.

For each of the six accretion prescriptions listed above, we consider a case where the background gravitational force $a_{{\rm bg},i}$ is always present and points inward. We then consider a second case, motivated in the previous subsection, where $a_{{\rm bg},i}$ is set to zero whenever the BH is more massive than the mass enclosed inside its present position [i.e. if $M_i > M_{\rm bg}(r_i)$].

Our full suite of simulations is summarized in Table 9. Each of the twelve models we have described above (six different accretion prescriptions, and two different treatments of the background gravitational force at small radii) were simulated multiple times using different initial values for BH masses, positions, and velocities (the criteria for our initial conditions are described in §4.3.2 and 8). We simulated each model using 43 distinct sets of initial conditions; each set of initial conditions was recycled 12 times, using the different prescriptions in the 12 different models.

4.7 Results

We now turn to the results of our N-body simulations. We briefly summarize our major findings below, and follow these with detailed explanations and analyses.

1. We found that in a majority (24 out of 43) of initial condition sets, an IMBH of mass $\sim 10^3$ to $\sim 10^5$ $M_\odot$ formed as a result of hyper-Eddington accretion, in all the models that allowed for this accretion mode (i.e. in models "$f_{\rm in} = 1$","$f_{\rm in} = 10^{-3}$","$f_{\rm in} = 0$" and "I" listed in Table 2).

2. If one set of initial conditions results in IMBH formation in one hyper-Eddington model, then it does so in all the others. The determining factor is whether the BH passes through a dense, gas-rich region (as a result of small semimajor axis, small pericenter, or both) where dissipation of orbital energy via gas drag is efficient.

3. All of the IMBHs end up within the central $\lesssim 0.01$ pc of the protogalactic halo, strongly suggesting that they are viable precursors of nuclear SMBHs observed as quasars at $z > 6$.

4. There is at most one IMBH in each simulation; we do not find a single instance of multiple IMBHs forming.

5. Many of the IMBHs capture a lighter BH into a close, sub-parsec orbit, and we argue that such systems could lead to EMRI events detectable by planned gravitational-wave observatories, such as *eLISA*.

6. We do not find mergers between stellar-mass BHs, in contrast to similar studies by Tagawa et al. (2015, 2016). The main reason for this appears to be simply the larger radii at which we initially place the BHs.

7. The above findings appear to be robust with respect to our treatment of both dynamical friction and the gravitational force due to the background matter distribution.

4.7.1 The onset of hyper-Eddington accretion

The first significant event in our simulations is the descent of the innermost BH to the dense gaseous core, which is driven by the decay of its orbit due to dynamical friction. Because the BHs are relatively widely separated in our initial conditions (the initial mean separations are $\simeq 75$ pc), three-body interactions at this stage are rare. As the innermost BH sinks even closer to the center, its ambient density increases and its Bondi accretion rate increases. The accretion rate eventually transitions from $\dot{M} = \dot{M}_{\rm B} < \dot{M}_{\rm Edd}/\eta$ (sub-Eddington Bondi), to $\dot{M} = \dot{M}_{\rm Edd}/\eta < \dot{M}_{\rm B}$ (Eddington-limited), and finally (in 25 out of 43 cases) to $\dot{M} = \dot{M}_{\rm B} > 3000\ \dot{M}_{\rm Edd}$ (hyper-Eddington; accretion unimpeded by radiation feedback), as dictated by Eq. (73). This qualitative picture is shared by all of our simulations in which hyper-Eddington accretion occurs.

This progression is illustrated in the left panel of Figure 20, which shows the journey in position and mass for a BH that undergoes hyper-Eddington accretion. The data is taken from a simulation run using the $f_{\rm in} = 1$ accretion scenario and with the background gravitational force always present—however, we reiterate that the behaviour shown here is shared by all examples of hyper-Eddington accretion in our simulations.

The top left panel of this figure shows the position of the innermost BH with the core radius $r_{\rm c}$ shown for reference. In the middle panel, we show the BH accretion rate in units of $\dot{M}_{\rm Edd}$, alongside the two critical values that determine the accretion regime according to Eq. (73): $3000\ \dot{M}_{\rm Edd}$, and $\dot{M}_{\rm in}$. Finally, in the bottom panel, we show the mass M of the BH (in $\rm M_\odot$). We also plot $M/M_{\rm B}$, the ratio of the BH mass to the gas mass inside its Bondi sphere. Recall that when this ratio is greater than unity, we study evolution under different accretion prescriptions (the difference between our $f_{\rm in} = 1$, $f_{\rm in} = 10^{-3}$, $f_{\rm in} = 0$, and "I" models; see Table 10). Further, we show $M/M_{\rm bg}$, the ratio of the BH mass to the mass contained in the halo inward of the BH's radial position. If this ratio is greater than unity, we also study BH evolution without the background gravitational in the "Background Potential Off" simulations (see Table 10).

For comparison, on the right side of the figure, we have plotted the same information for another BH in the same simulation that does *not* undergo hyper-Eddington accretion. This BH is on a highly elliptical orbit around the center of the halo, but its Bondi accretion value never exceeds the $\dot{M}_{\rm B} = 3000\ \dot{M}_{\rm Edd}$ threshold to overcome radiative feedback.

In the left side of this figure, we see that it takes the accreting BH only a few Myr to sink to the center where it begins to undergo hyper-Eddington accretion. For BHs that undergo hyper-Eddington accretion in our simulations that allows this (models $f_{\rm in} = 1$, $f_{\rm in} = 10^{-3}$, $f_{\rm in} = 0$, and "I"), an average of 5.6 Myr elapses between when the Bondi accretion rate first reaches the Eddington limit and when it reaches $3000\dot{M}_{\rm Edd}$. This is much shorter than the Salpeter time (~ 45 Myr for the adopted radiative efficiency $\eta = 0.1$), indicating

that the increase in the Bondi accretion rate is caused by the increase in ambient density ($\dot{M}_{\rm B} \propto \rho_{\rm gas} \propto r^{-2}$ for $r \gtrsim r_{\rm c}$), as opposed to mass growth ($\dot{M}_{\rm B} \propto M^2$).

In Figure 21, we show the semimajor axis (vertical axis) and eccentricity (plotted as $1-e$, horizontal axis) of each of our BHs in the "$f_{\rm in}=1$, $a_{\rm bg}$ on" simulation set (a total of 43 runs and 430 BHs, 24 of which grow to become IMBHs). Because the BH orbits are not Keplerian (and therefore not elliptical), the concept of eccentricity is not rigorously defined. We evaluate the instantaneous eccentricity using the standard formula for Keplerian orbits

$$e = \sqrt{1 + \frac{2\epsilon\ell^2}{\{G[M+M_{\rm bg}(r)]\}^2}}, \tag{74}$$

where ϵ is the specific energy of the orbit (orbital energy divided by the "instantaneous reduced mass" $M\ M_{\rm bg}(r)/[M+M_{\rm bg}(r)]$), and ℓ is the specific angular momentum (orbital angular momentum divided by the instantaneous reduced mass). Note that the gravitational potential used in calculating ϵ is logarithmic, and the enclosed central mass $M_{\rm bg}(r)$ varies with the orbital radius.

The red dots in Figure 21 represent the orbital evolution of BHs that do not make it to the dense central region to become IMBHs within the 500 Myr runtime of the simulations; the black curves represent the orbital evolution of BHs that do grow into IMBHs. The dotted box in the upper left approximately encloses initial orbital parameters (a and e) for BHs which do not grow to IMBH/SMBH. BHs with the initial a and e outside the box have sunk to the core and have experienced hyper-Eddington accretion, whereas we find no such examples with initial a and e that lie inside the box. Notice that some BHs with small initial semimajor axes (i.e. outside the box) migrate to the center after the first BH becomes massive, and form a bound pair with it. The dashed horizontal line near the bottom marks the core radius, $r_{\rm c} = 0.003$ pc.

BHs evolve from having large semimajor axes and high-eccentricity orbits (upper left of the panel) about the center of the protogalaxy potential, to having tighter, nearly circular orbits (lower right portion of the panel). As a BH approaches the center of the protogalaxy, it sinks to the center more quickly than it can complete an orbit. The final plunge into the center of the protogalaxy is not plotted, as we find that e cannot be reliably calculated from the shape of the orbit.

The orbital evolution and the final transition to hyper-Eddington accretion described above can be understood as follows. Outside the core, the mass enclosed inside the BH orbit is $M_{\rm bg} \approx 4\pi\rho_{\rm c} r_{\rm c}^2 r \propto r$, and the dynamical time can be expressed as

$$\tau_{\rm dyn} = \left(\frac{r}{a_{\rm g}}\right)^{1/2} \approx 170 \left(\frac{r}{r_{\rm c}}\right)\ {\rm yr}\,. \tag{75}$$

Barring an encounter with another BH, BHs have velocities $v \sim c_{\rm s}$ or smaller. In this

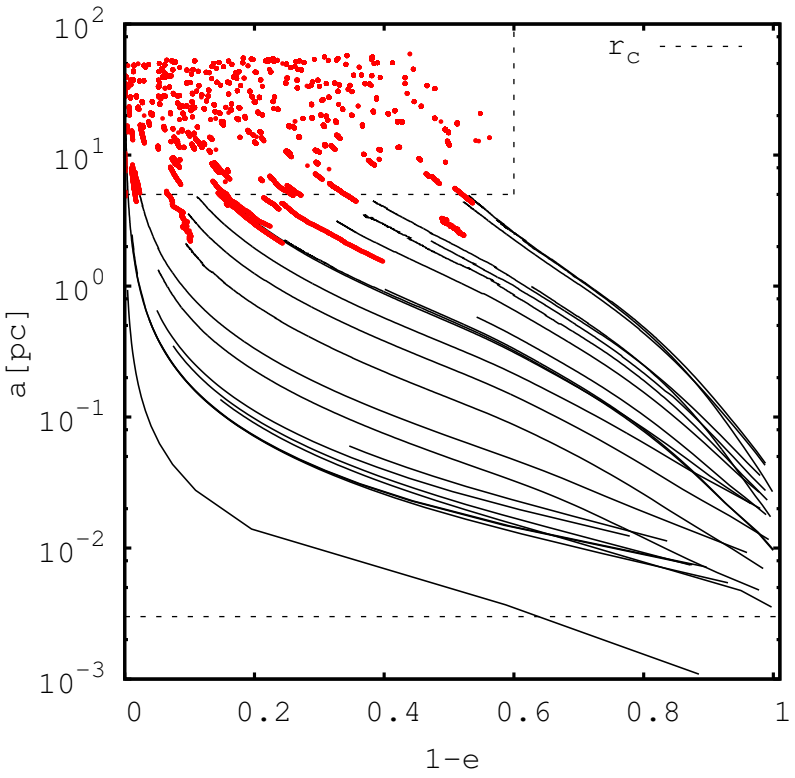

Figure 21: The evolution of the semimajor axis and eccentricity for 430 BHs in 43 simulations with the prescription "$f_{\rm in} = 1$, $a_{\rm bg}$ on." The BHs begin with large semimajor axes and high-eccentricity orbits (upper left of the panel). The red dots represent the orbits of 406 BHs that do not sink to the center within the simulation runtime of 500 Myr. The black lines represent the orbits of the 24 BHs that sink to the dense central region and grow into IMBHs. The orbits circularize as they decay, before finally plunging radially to the center (this last phase is not shown in the figure). Only BHs with an initial distance of $\lesssim$ 5pc from the center are found in this category. The dotted box region on the upper left demarcates the region in the a-e parameter space where we never found examples of BHs that successfully grow into a central massive BH. The dashed horizontal line marks the core radius at $r_{\rm c} = 0.003$ pc.

subsonic regime, $f^{\rm gas}v^{-3} \lesssim c_{\rm s}^{-3}$ in Eq. 65. Writing $a_{\rm df} \sim v/\tau_{\rm df}$, we can estimate $\tau_{\rm df}$ as

$$\tau_{\rm df} \sim (1-4) \times 10^3 \left(\frac{M}{10\ {\rm M}_\odot}\right)^{-1} \max\left[1, \left(\frac{r}{r_{\rm c}}\right)^2\right]\ {\rm yr}\,, \tag{76}$$

where the leading factor depends on whether the Mach number is less than or greater than $\mathcal{M} = 0.8$. The azimuthal force due to dynamical friction results in circularization and orbital decay on a timescale of $\tau_{\rm df}$, whereas in the radial direction the background force dominates over dynamical friction ($\sim$ Myr for pericenter passage of $\gtrsim 30\ r_{\rm c}$ and $M \sim 30\ {\rm M}_\odot$). Because $\tau_{\rm df} \propto r^2$ outside the core, the decay accelerates, with the final stages occurring over a few thousand years.

Since the orbital decay occurs on a timescale much shorter than the Salpeter time, the BH does not grow significantly by either sub-Eddington or Eddington-limited accretion. If $M \lesssim 60\ {\rm M}_\odot$, then $r_{\rm B} \lesssim r_{\rm c}$ and the orbital velocity is comparable to or less than the sound speed. The Bondi accretion rate for stellar-mass BHs outside the core can be estimated as

$$\begin{aligned}
\dot{M}_{\rm B} &\approx \frac{4\pi G^2 M^2 \rho(r)}{c_{\rm s}^3} \\
&= 7 \times 10^{-3} \left(\frac{M}{10\ {\rm M}_\odot}\right)^2 \left(\frac{r}{r_{\rm c}}\right)^{-2}\ {\rm M}_\odot\ {\rm yr}^{-1} \\
&= 3.1 \times 10^5 \left(\frac{M}{10\ {\rm M}_\odot}\right) \left(\frac{r}{r_{\rm c}}\right)^{-2} \dot{M}_{\rm Edd}.
\end{aligned} \tag{77}$$

Hence an infalling stellar-mass BH begins to undergo hyper-Eddington accretion when $r \gtrsim r_{\rm c}$.

4.7.2 Hyper-Eddington accretion: a brief but dramatic growth spurt

Once the BH overcomes the $\dot{M}_{\rm B} = 3000\ \dot{M}_{\rm Edd}$ threshold, its accretion rate instantaneously increases by a factor of $3000\ \eta = 300$. At the instant after this transition, its mass growth timescale is $M/\dot{M} \approx 0.1$ Myr (the Salpeter time divided by 3000η), but then rapidly shortens as $\propto M^{-1}$ before quickly hitting the ceiling $f_{\rm in}\dot{M}_{\rm in}$ imposed by the large-scale mass inflow rate (eq. 72). For comparison, note that in Eddington-limited growth, $\dot{M} \propto M$ and the growth timescale is constant at the Salpeter value. As a result, when the BH accretion rate becomes hyper-Eddington, its mass shoots up from a few $\times\ 10\ {\rm M}_\odot$ to more than $10^3\ {\rm M}_\odot$ in less than ~ 0.1 Myr.

This rapid mass growth results in two significant transitions in our simulations. First, the gas mass inside the Bondi radius exceeds that of the BH mass. Again, how the BH is assumed to accrete mass when $M_{\rm B} > M$ is what distinguishes our $f_{\rm in} = 1$, $f_{\rm in} = 10^{-3}$, $f_{\rm in} = 0$ and "I" models.

If $M \lesssim 60\ {\rm M}_\odot$, then the Bondi radius is small and the density at the surface of the Bondi sphere is close to the local density. Then, considering the Bondi-like profile ($\sim r^{-3/2}$) inside

the Bondi sphere,

$$M_{\rm B} \sim \int_0^{r_{\rm B}} \langle \rho_{\rm B}(r') \rangle \left(\frac{r_{\rm B}}{r'}\right)^{3/2} \, r'^2 \, {\rm d}r' \tag{78}$$

$$\sim \frac{8\pi}{3} r_{\rm B}^3 \rho(r)$$

$$\sim 0.7 \, {\rm M}_\odot \left(\frac{M}{10 \, {\rm M}_\odot}\right)^3 \min\left[1, \left(\frac{r}{r_{\rm c}}\right)^{-2}\right], \tag{79}$$

and $M_{\rm B} < M$.

However, since $r_{\rm B} \propto M$, at larger BH masses the Bondi sphere quickly becomes larger than the gaseous core. We just established above that the BH typically begins its hyper-Eddington accretion near the core. Therefore, once a BH grows to $M \gtrsim 60 \, {\rm M}_\odot$, $r_{\rm B} > r_{\rm c}$ and the gas density at the surface of the Bondi sphere is essentially given by the halo gas profile evaluated at $r = r_{\rm B}$. Then we can write

$$M_{\rm B} \sim \frac{8\pi}{3} r_{\rm B}^3 \rho(r_{\rm B}) \approx \frac{8\pi}{3} \rho_{\rm c} r_{\rm c}^2 r_{\rm B} = 2.4M > M. \tag{80}$$

We conclude that as soon as hyper-Eddington accretion begins, the gas enclosed inside the Bondi sphere shoots above the BH mass.

The second transition that occurs as the BH grows is that the BH mass exceeds the mass of the matter enclosed inside its orbit around the center of the halo ($M_{\rm bg}$). For $r \gg r_{\rm c}$, the enclosed mass is simply

$$M_{\rm bg} \gtrsim 4\pi \rho_{\rm c} r_{\rm c}^2 r \approx 200 \, {\rm M}_\odot \left(\frac{r}{r_{\rm c}}\right), \tag{81}$$

whereas inside the core ($r \ll r_{\rm c}$)

$$M_{\rm bg} = \frac{4\pi}{3} \rho_{\rm c} r^3 \approx 70 \, {\rm M}_\odot \left(\frac{r}{r_{\rm c}}\right)^3. \tag{82}$$

Either way, as the BH grows beyond several 100 $\rm M_\odot$ near or inside the core, our simulations always result in $M > M_{\rm bg}$. Since the BH dominates the central potential, the innermost gas distribution will be strongly disturbed, and will correspond to a radial force towards the center of the halo. This is the motivation for running a second set of simulations, in which the gravitational force of the background matter is turned off if $M > M_{\rm bg}$.

The above caveats aside, the hyper-Eddington rate initially follows the Bondi rate and scales as $\dot{M} \propto M^2$. One significant aspect of this mode of growth is that, in addition to the raw accretion rate being much higher than Eddington-limited growth, the accretion timescale decreases (i.e. the growth rate accelerates and is faster than exponential). However, this growth does not last long in our simulations, because $\dot{M}$ encounters one of two upper limits.

The first upper limit is $f_{\rm in}\dot{M}_{\rm in}$ (where $\dot{M}_{\rm in} = c_{\rm s}^3/G$), the parametrized gas supply rate into the center of the halo in the "$f_{\rm in} = 1$","$f_{\rm in} = 10^{-3}$", and "$f_{\rm in} = 0$" models.

The second one is due to the fact that as M increases, $r_{\rm B}$ increases, and the density at the surface of the Bondi sphere decreases. For $r_{\rm B} \gg r$ and $r_{\rm B} \gg r_{\rm c}$, the density at the Bondi

Table 10: Average mass, mass ratio, eccentricity (e) and accretion rate for the first BH-BH pair when its semimajor axis is $a \approx 1$ pc. Accretion rates are in units of the Eddington rate. The subscript "1" refers to the more massive BH, and "2" to the less massive BH.

$a_{\rm bg}$	Model	$M_1[\,{\rm M}_\odot]$	$M_2[\,{\rm M}_\odot]$	$q\ (=M_2/M_1)$	e	$\dot{M}_1/\dot{M}_{\rm Edd,1}$	$\dot{M}_2/\dot{M}_{\rm Edd,2}$
on	$f_{\rm in}=0$	160	71	0.45	0.92	0	38000
	$f_{\rm in}=10^{-3}$	1.5×10^4	45	3.0×10^{-3}	0.99	19	4.8
	$f_{\rm in}=1$	6.1×10^5	31	5.1×10^{-5}	0.90	760	8.6×10^{-2}
	I	2.5×10^5	38	1.5×10^{-4}	0.87	10	2.9
	E	450	50	0.11	0.99	10	5.9
	$\dot{M}=0$	41	37	0.92	0.96	0	0
off	$f_{\rm in}=0$	170	91	0.55	0.92	0	58000
	$f_{\rm in}=10^{-3}$	2.6×10^4	44	1.7×10^{-3}	0.99	17	3.2
	$f_{\rm in}=1$	4.1×10^6	29	7.0×10^{-6}	0.84	240	2.1×10^{-2}
	I	2.7×10^6	33	1.2×10^{-5}	0.98	10	0.78
	E	370	53	0.14	0.90	10	6.8
	$\dot{M}=0$	41	37	0.92	0.87	0	0

sphere surface becomes

$$\rho(r_{\rm B}) \approx \rho_{\rm c}\left(\frac{r_{\rm B}}{r_{\rm c}}\right)^{-2} = \frac{\rho_{\rm c} r_{\rm c}^2 c_{\rm s}^4}{4G^2M^2}. \tag{83}$$

Then the Bondi accretion rate evaluates to

$$\dot{M}_{\rm B} = \frac{4\pi G^2 M^2 \rho(r_{\rm B})}{c_{\rm s}^3(1+\mathcal{M}^2)^{3/2}} \approx \pi\rho_{\rm c} r_{\rm c}^2 c_{\rm s} \approxeq 0.2\ {\rm M}_\odot\ {\rm yr}^{-1}, \tag{84}$$

or ≈ 40 % of $\dot{M}_{\rm in}$ $(=c_{\rm s}/G^3)$.

To summarize, the dynamics in our simulations evolves according to the following trends:

1. The orbit of the innermost BH decays on the dynamical friction timescale. As it does, the accretion rate goes from sub-Eddington Bondi-Hoyle-Littleton ($\dot{M} \propto M^2$, $\dot{M} \leq \dot{M}_{\rm Edd}/\eta$) to Eddington-limited accretion ($\dot{M} = \dot{M}_{\rm Edd}/\eta \propto M$). However, this phase lasts much less than a Salpeter time, and the mass growth is typically insignificant.

2. As the BH approaches the center of the halo, typically at $r \sim$ a few $\times\, r_{\rm c}$, the Bondi rate $\dot{M}_{\rm B} > 3000\dot{M}_{\rm Edd}$. At this point, following Inayoshi et al. (2016), we assume that photon trapping allows for hyper-Eddington accretion, i.e. once again matching the unimpeded BHL rate ($\dot{M} = \dot{M}_{\rm B} \propto M^2$).

3. In the hyper-Eddington phase, the BH grows from a typical mass of a few $\times 10$ $M_\odot$ to $\gtrsim 10^3$ $M_\odot$. During this rapid transition, the BH becomes more massive than the mass contained inside its orbit around the halo center ($M > M_{\rm bg}$), and the gas mass enclosed within its Bondi sphere exceeds its own mass ($M_{\rm B} > M$). The behaviour up to this point is almost identical for all the simulations that allow for hyper-Eddington accretion (models "$f_{\rm in} = 1$","$f_{\rm in} = 10^{-3}$","$f_{\rm in} = 0$" and "I"). Most of the variation between these models result from the difference in prescriptions when $M_{\rm B} > M$ and $M > M_{\rm bg}$. That is, the models "branch out" from this point forward.

4. The accretion rate then slows, as it encounters the halo mass supply limit $f_{\rm in}\dot{M}_{\rm in}$ or the asymptotic constant value of $\dot{M}_{\rm B}$ in the limit of large Bondi radius. Because both of these values are constant, as the BH mass increases $\dot{M}$ falls below $3000\dot{M}_{\rm Edd} \propto M$, and becomes Eddington-limited again. The final masses of the BHs, and their configuration with respect to the halo and other BHs, depend on the prescriptions for accretion and background gravitational force, as discussed in the following.

4.7.3 The final IMBH masses and configurations

In Table 10, we summarize the average final masses and accretion rates found in our simulations, for each combination of our prescriptions for gas accretion and treatment of the background gravitational force. The values presented are the mean values over the 24 realizations per model with a nuclear BH, evaluated when the first BH has sunk to the center of the protogalaxy and has captured a companion BH into a closed orbit with a semimajor axis ≈ 1 pc. We have chosen to stop the simulations at $a = 1$ pc, because at smaller separations between the innermost bound pair three-body scatterings are rare. We assume that past this separation, the inner bound pairs evolve through damped, closed orbits until merger. masses and instantaneous accretion rates of the central BH are denoted with a subscript "1," and those for the smaller companion BH with a subscript "2." We also list the mass ratio $M_2/M_1 \leq 1$ and the orbital eccentricity e of the pair.

The bottom half of Table 10 lists the BH properties found in simulations where the background gravitational force exerted on a given BH was set to zero whenever the BH mass exceeded the mass enclosed inside its radial position. The final values found in these simulations do not vary significantly from the ones in which the background force was always present (the top half of the table).

4.7.4 The central BH

Let us first discuss the central BH. In the simulations with $f_{\rm in} = 0$, BHs stop growing once they are outweighed by the gas mass enclosed inside their Bondi sphere. Because of this, they never grow beyond $M_1 \sim 100$ $M_\odot$. In simulations where $f_{\rm in} = 10^{-3}$, $f_{\rm in} = 1$, and "I" the growth rate is capped by $f_{\rm in}\dot{M}_{\rm in}$ and by the asymptotic Bondi rate (eq. 84). These upper bounds allow the growth of the central BH to $M_1 \sim 10^4$ $M_\odot$ for $f_{\rm in} = 10^{-3}$, and $M_1 \sim 10^5$ $M_\odot$ to $\sim 10^6$ $M_\odot$ for $f_{\rm in} = 1$ and "I" (note that the values for $f_{\rm in}\dot{M}_{\rm in}$ and the asymptotic Bondi rate are comparable). *In all simulations where hyper-Eddington accretion is allowed to continue past the point $M < M_{\rm B}$, the central BH grows into an IMBH or SMBH.*

For reference, we can see that if the mass growth is limited at the canonical Eddington value (model "E"), then the central BH which forms a bound pair with another BH does not grow significantly. This is because the bound pairs reach $a \lesssim 1$ pc soon after the first BH sinks to the core, when both BHs are still close to their original stellar masses. Once a tight central binary forms, its orbital velocity increases, and suppresses the BHL accretion rate below the Eddington value, stunting further growth of either BH. We also show, for reference, values in which accretion is not allowed at all (model "$\dot{M} = 0$").

4.7.5 The stellar-mass BH companion

A striking result of our simulations is that we find no more than one hyper-accreting BH per simulation (i.e. either zero or one IMBH per galaxy). The reason for this is that if an IMBH forms in one of our simulations, it prevents other BHs from undergoing hyper-Eddington accretion.

This finding can be explained as follows. The first IMBH forms relatively quickly (its orbit decays in a few Myr), and does so at the center of the halo, where gas densities (and Bondi accretion rates) are high. Once this IMBH is in the center of the halo, any subsequent BH whose orbit decays will be captured by the IMBH potential. Whereas the first BH had orbital velocities $v \lesssim c_{\rm s}$ as it fell toward the center, the orbital velocity of a BH captured by the IMBH will be supersonic. The supersonic orbital motion suppresses both the orbital decay rate via dynamical friction, preventing the second BH from sinking deep into the gas-rich center of the halo. On top of this, the high velocity suppresses the Bondi-Hoyle-Lyttleton accretion rate, $\dot{M}_{\rm B} \propto (c_{\rm s}^2 + v^2)^{-3}$.

Thus, the first BH to wander to the center of the potential is able to grow to an IMBH via hyper-Eddington accretion, but then *subsequently prevents other BHs from doing the same.* As a result, our simulations typically produce bound pairs of IMBHs and stellar-mass BHs with mass ratios $q \equiv M_1/M_2 \sim 10^{-4} - 10^{-2}$. Such "extreme mass-ratio" systems are one of the important targets for detection by planned gravitational-wave instruments, and we will revisit them in our discussion section. As Table 10 shows, interestingly, all EMRIs have a highly eccentric orbit; this is a result of the preferential capture of stellar-mass BHs on such orbits, i.e. with pericenters inside the sphere of influence of the newly grown IMBH.

In models where an IMBH is not produced, subsequent BHs are free to fall to the center at low speeds, just as the first one did. As shown in Table 10, in the "$f_{\rm in} = 0$" models, we find near-equal, stellar-mass binaries forming in the nucleus. In these models, the growth of the 1st BH is artificially stunted, allowing the 2nd BH to experience a brief phase of hyper-Eddington accretion. This hyper-Eddington phase, as the 2nd BH travels through the dense gaseous core, only lasts until it, too, reaches a mass of $M_2 \gtrsim 100$ M$_\odot$; its Bondi sphere will then become self-gravitating, and its grow is terminated, just as for the first BH.

4.7.6 Different treatments of central background potential

The main difference in the two sets of models (shown in the bottom vs. top half of Table 10) is that if $a_{\rm bg}$ is always present, the central BH ends up at the very center of the model protogalaxy, because the background force always continues to point inward. In contrast, when $a_{\rm bg}$ is turned off, the gas drag brings the BHs to rest near—but not *at*—the center.

Aside from this detail, we find no major qualitative difference in the properties of the BHs across these two sets of simulations. We conclude that the hydrodynamical reaction of the innermost gas to the central BH is unlikely to significantly influence our main conclusions about the demography and location of the emerging BH population (in particular whether hyper-Eddington accretion occurs).

4.8 Discussion

4.8.1 SMBH precursors

Our simulations focus on the growth and orbital evolution of Pop III remnant BHs in a model protogalaxy that is just above the atomic-cooling threshold for virial mass. We find that the hyper-Eddington accretion prescription of Inayoshi et al. (2016) typically results in the formation of a single IMBH in the center of the protogalaxy.

The natural interpretation is that this is a massive nuclear BH that will continue to grow as the host galaxy grows. The formation of a $\sim (10^4 - 10^5)$ $M_\odot$ BH in an atomic-cooling halo is the same end result as in the so-called "direct collapse BH" scenario. These IMBHs must then grow at a logarithmically time-averaged rate $\dot{M} \lesssim 10\dot{M}_{\rm Edd}$ (i.e. at a rate comparable to the Eddington limit for a radiative efficiency $\eta \sim 0.1$) to explain the $\gtrsim 10^9$ $M_\odot$ engines of the $z \gtrsim 6$ quasars. The direct collapse scenarios generically require specific conditions that may be rare in the Universe. For example, in most models the collapse is facilitated by a high Lyman-Werner intensity that dissociates molecular hydrogen, and thus only a small fraction of galaxies are expected to be viable direct-collapse sites (e.g. Dijkstra et al. 2008; Shang et al. 2010; Hosokawa et al. 2012; Dijkstra et al. 2014; Sugimura et al. 2014; Visbal et al. 2014; Latif et al. 2015; Inayoshi & Tanaka 2015; Regan et al. 2016).

The picture suggested by our simulations is that IMBHs in protogalactic nuclei could plausibly be more commonly produced by hyper-Eddington growth of a pre-existing stellar-mass BH – the essential requirements being only a high-density core, and a large-scale inflow rate of $O(\ M_\odot {\rm yr}^{-1})$, down to the Bondi radius ($\sim$0.01 pc) of the central BH with initial mass of ~ 100 $M_\odot$.

These two hypotheses for SMBH progenitors—rare direct-collapse seeds and more common results of hyper-Eddington accretion—could be tested against observations through event rates detected by milli-Hertz gravitational-wave detectors (e.g. Sesana et al. 2007; Tanaka & Haiman 2009) or by the global signatures of the redshifted 21 cm line (e.g. Tanaka et al. 2016).

4.8.2 EMRI detections

Whenever an IMBH forms in our simulations, we find that it captures one or more stellar-mass BH companions. Mergers of such BHs are predicted to produce EMRIs, a category of gravitational wave events that is one of the primary *low-redshift* targets of the space-based interferometer *eLISA* (Consortium et al., 2013; Amaro-Seoane et al., 2013).

Because the timescale for the $M_2 \ll M_1$ pairs in our simulations to merge through emission of gravitational waves is well over a Hubble time, additional mechanisms such as three-body scatterings or continuous gaseous dissipation (e.g. by a circumbinary accretion

disc, Cuadra et al. 2009; Roedig et al. 2011) are required to drive the merger. The production of such pairs in our simulations suggests that they could result in mergers of IMBHs and stellar-mass BHs at lower redshifts.

We note that our IMBH-BH pairs have eccentricities $e \gtrsim 0.9$, at $a \approx 1$ pc (Table 10). This points to the interesting possibility that they could lead to EMRIs that have residual eccentricities when they enter the eLISA band. However, as we do not follow the evolution of such pairs all the way to merger, and given the variety of possibly relevant mechanisms, we leave the assessment of any residual eccentricity in the eLISA band for future work.

Additionally, the merger of a protogalaxy or dwarf galaxy containing an IMBH with a more massive one containing a SMBH should result in the formation of a SMBH-IMBH pair. While this is an expected corollary of our findings, further work is required to assess whether such pairs can overcome the so-called "final parsec problem" (Merritt & Milosavljević, 2005).

4.8.3 Comparison with previous work

Several recent papers have investigated super-Eddington accretion in galaxies. Lupi et al. (2016) used hydrodynamical simulations to investigate super-Eddington accretion of stellar-mass BHs in circumnuclear gas discs in the hearts of galaxies. Their scheme allows gas particles close to the BH to accrete and grow the BH. Pezzulli et al. (2016) considered the growth of stellar-mass BHs inside model galaxies that account for metal enrichment, dust, star formation and detailed cooling. They assume that the central BH grows at a rate proportional to the cold gas mass in the bulge, and inversely proportional to the dynamical time of the bulge.

One major difference between this study and those papers is the accretion prescription. We adopt the analytic Bondi-like accretion prescription based on Inayoshi et al. (2016) and featuring the transition from low Bondi-Hoyle-Lyttleton accretion, to Eddington-limited accretion, and then to hyper-Eddington accretion. Another notable difference is that Lupi et al. (2016) and Pezzulli et al. (2016) examined the growth of BHs in fully evolved galaxies, whereas in this study we focus on the growth of Pop III remnant BHs into IMBHs in a protogalaxy of mass $\sim 10^8$ $M_\odot$.

Alexander & Natarajan (2014) considered a set-up similar to ours, in which a stellar-mass BH is in orbit in a protogalaxy, accreting above the fiducial Eddington rate. The focus of that paper was to assess the ability of the gas inside the BH's sphere of influence to shed angular momentum and accrete onto the BH. The orbit of the BH was assumed to be determined by its interactions with a nuclear star cluster (which was found to be important to reduce the angular momentum). Here we treated a small system of BHs, and assumed that the background gas dominates their orbital decay into the nucleus.

Closest to our study, Tagawa et al. (2015) and Tagawa et al. (2016) performed N-body simulations of BHs embedded in a compact distribution of gas, in order to gauge the merger mechanisms of BHs in galactic centers. Overall, the set-up and goals of these studies and ours are similar, although Tagawa et al.'s focus was to clarify the occurrence rate and mechanisms of stellar-mass BH mergers. Our most notable finding – the frequent formation of a *single IMBH* at the *center* of the protogalaxy – differs from the conclusions by Tagawa et al. (2015) and Tagawa et al. (2016), who find efficient formation of stellar-mass binaries, often facilitated by 3-body interactions.

These differences in conclusions arise from three important differences between our initial conditions and model assumptions. First, we spread 10 initial BHs over a large region of up to 100 pc, with separations of > 10 pc. The initial BH separations are much more compact in (Tagawa et al., 2015, $0.01 - 10$ pc) and especially in (Tagawa et al., 2016, $0.01 - 0.1$ pc). As a result, we do not find 3-body interactions or stellar-mass binaries. Second, we adopt a centrally condensed density profile, while Tagawa et al. (2015) and Tagawa et al. (2016) both assume homogeneous clouds. Third, we investigate various accretion prescriptions, and allow hyper-Eddington accretion at rates limited only by the steady large-scale inflow rate (whereas Tagawa et al. 2015 did not consider accretion onto BHs and Tagawa et al. 2016 considered an accretion rate capped by the Eddington limit and by the assumed total cloud mass). As a result of these last two differences, we find that rapid growth into IMBHs and SMBHs is much more common, and always occurs in the nucleus – producing EMRIs, rather than stellar-mass binaries.

4.8.4 Caveats

Our results were obtained in simplified toy models, and are subject to several caveats. We here discuss three possible major limitations of our model.

Gravitational potential of the background matter. Our simulations assume a static density profile of gas and dark matter, instead of allowing the protogalaxy to evolve dynamically or thermodynamically. On one hand, the simplified treatment of the protogalaxy allowed us to run hundreds of simulations—dozens of initial condition realizations for each of a dozen different combinations of theoretical models. On the other, this survey of model prescriptions and the statistical sample size came at the expense of a more detailed treatment of gas dynamics.

In particular, not accounting for the dynamical evolution of the surrounding matter directly impacts the two gravitational effects in this work: the gravitational force exerted by the ambient matter, and dynamical friction. Our assumption of a static background allows us to evaluate analytically the gravitational force from the ambient medium, and we treat the dynamical friction using a modified version of the Chandrasekhar formula Binney & Tremaine 1987.

In an attempt to gauge the robustness of our results with respect to these theoretical simplifications, we ran different sets of simulations with very different assumptions. First, as discussed above, we ran a set of simulations in which the gravitational field of the background material was entirely removed, once the BH mass exceeded the mass of the matter enclosed inside its orbit. The only qualitative difference we found was that this resulted in the central BH ending up slightly off-center (whereas leaving the analytic background force "on" had the effect of always pulling the central BH to the very center of our model protogalaxy). We also note that while gas anisotropies are common features in simulations of protogalactic haloes, the masses of any anisotropic clumps tend to be small compared to that of the ambient gas (which are overall well-described by power-law profiles, e.g. Regan & Haehnelt 2009).

We also ran, for the same set of the different accretion prescriptions, simulations in which the dynamical friction was not calculated using the local gas density at the BH coordinates $\rho(r)$, but using the gas density averaged over the surface of its Bondi sphere, $\langle \rho_{\rm B}(r) \rangle$. The

rationale behind this experiment was that whereas the derivation of the Chandrasekhar formula assumes an infinite, uniform background distribution of gas, the physical phenomenon of dynamical friction is due to the wake of overdense gas formed at some distance from the massive body. We found no qualitative difference between this set of simulations and the one described in §3. While there is a rich literature on quantifying how dynamical friction differs in nonuniform density distributions or nonlinear trajectories (e.g Sánchez-Salcedo & Brandenburg, 2001; Just & Peñarrubia, 2005; H. Kim & Kim, 2007; W.-T. Kim, 2010), these studies do not report major differences from the Chandrasekhar formula.

We conclude that our results are unlikely to be an artifact of the simplified treatment of the gravity of the ambient matter distribution.

Star formation in the halo. Our model does not consider star formation inside the halo. Once a dense region of gas cloud in the halo becomes optically thick, fragmentation leads to star formation (e.g. Regan et al., 2014; Becerra et al., 2015, and refs. therein). The stars may form directly in the core region, or form elsewhere and subsequently migrate to the core. Because the stars have lifetimes of $\sim$ Myr, we expect them to become BHs before arriving at the core, or shortly afterward. In practice, we do not expect significant qualitative differences based on which type of BH reaches the core first—the pre-existing BHs our simulations had in mind, or the stars/BHs that form in situ in the halo. According to the picture suggested by our simulations, whichever type of BH arrives at the center first should grow massive via hyper-Eddington accretion, then capture any subsequent arrivals into bound orbits (which may eventually produce EMRI events).

Therefore, we do not expect any in situ star formation in the halo to qualitatively affect our findings.

Effect of radiation on the hydrodynamics. Our toy model neglects the radiation produced by the accreting BHs (and of any stars found in the same galaxy). The gas in the halo cools efficiently and is mostly neutral, and accreting BHs will create their individual small HII regions. Here we estimate the size of these HII regions, before the BHs wander into the core and reach the hyper-Eddington state. In the "intermediate" regime, when $\eta^{-1} \lesssim \dot{M}_{\rm B}/\dot{M}_{\rm Edd} \lesssim 3000$, the time-averaged accretion rate is limited to the Eddington rate, because of the periodic formation, disappearance, and re-appearance, of an HII region that makes the accretion episodic. The maximum size of this HII region is larger than the Bondi radius, by definition, before the hyper-Eddington state can be reached Inayoshi et al. (2016). Assuming a luminosity of $L_{\rm Edd}$, the HII region size is $R_{\rm HII} = 8 \times 10^{13} (M/30\ {\rm M}_\odot)^{1/3} (r/r_c)^{4/3}$cm (see eq. 27 in Inayoshi et al. 2016) or $R_{\rm HII}/r = 0.01 (M/30\ {\rm M}_\odot)^{1/3} (r/r_c)^{1/3}$. This means that the HII region remains relatively small for BHs located within a few pc of the core. A near-Eddington BH outside this region (say at 10-100 pc) could blow a large HII bubble and change the global density distribution. However, the stellar-mass BHs in these outer, low-density regions will be highly sub-Eddington and dim (eq. 77). We conclude that radiative feedback is unlikely to prevent the onset of the hyper-Eddington phase of the first BH that sinks to the central region.

Validity of Hyper-Eddington accretion. As emphasized throughout this paper, a key ingredient of our model is that we allow rapid accretion, well in excess of $L_{\rm Edd}/c^2$. This is based

on the recent results of Inayoshi et al. (2016), who find this to be the case for accreting BHs whose HII region is confined to within the Bondi radius. This conclusion is subject to a few caveats summarized in Inayoshi et al. (2016). In particular, here we highlight the fact that Inayoshi et al. (2016) assumes a spherically symmetric accretion flow with low angular momentum, such that the centrifugal radius (setting the size of an accretion disc, producing significant luminosity) is smaller than the trapping radius (inside which photons are advected inward with the flow, rather than diffusing out). At the onset of the hyper-Eddington phase, the latter is 1500 $R_{\rm Sch}$, or $5 \times 10^{10} (M/100\ {\rm M}_\odot)$ cm. We thus require that the accretion flow onto the BHs remain quasi-spherical down to this distance from the BH. While this appears feasible once the BH settles to the bottom of the potential well, a proper assessment will require follow-up investigations, resolving the angular momentum transfer and dissipation for the flow onto the central BH. However, we note here that this requirement is much less stringent than the one addressed in a similar context by Alexander & Natarajan (2014), who required the centrifugal radius to be as small as a few $\times R_{\rm Sch}$ (and found it to be feasible, faciliated in their model by resonant effects due to a central stellar cluster).

4.9 Summary

In this paper, we described the formation of 10^{3-5} $\rm M_\odot$ IMBHs in centrally condensed gas clouds, arising from a small cluster of Population III remnant BHs. The stellar-mass BH are assumed to have been delivered into the cloud during the process of the hierarchical assembly of the halo via mergers, and have a spatially extended initial configuration (several pc to 100 pc). We then follow their accretion and orbital dynamics via an N-body calculation. These calculations reveal that as a result of gas drag, one of these BHs typically sinks to the nucleus, where it rapidly grows into an IMBH.

Our results suggest a viable pathway to forming the earliest massive BHs in the centers of early galaxies. We also find that only one IMBH can form in this way per galaxy, and that this IMBH typically captures a stellar-mass BH companion, making these systems observable in gravitational waves as EMRIs with *eLISA*.

More detailed simulations that account for the hydrodynamics, radiative transfer, and the cosmological evolution of the host protogalaxy are required to test this idea further.

5 Interactions between multiple supermassive black holes in galactic nuclei: a solution to the final parsec problem

T. Ryu, R. Perna, Z. Haiman, J. Ostriker, N. Stone *The Monthly Notices of the Royal Astronomical Society*, Vol. 473, 3410 (2018)

Abstract

Using few-body simulations, we investigate the evolution of supermassive black holes (SMBHs) in galaxies ($M_\star = 10^{10} - 10^{12}$ M$_\odot$ at $z = 0$) at $0 < z < 4$. Following galaxy merger trees from the Millennium simulation, we model BH mergers with two extreme binary decay scenarios for the 'hard binary' stage: a full or an empty loss cone. These two models should bracket the true evolution, and allow us to separately explore the role of dynamical friction and that of multi-body BH interactions on BH mergers. Using the computed merger rates, we infer the stochastic gravitational wave background (GWB). Our dynamical approach is a first attempt to study the dynamical evolution of multiple SMBHs in the host galaxies undergoing mergers with various mass ratios ($10^{-4} < q_\star < 1$). Our main result demonstrates that SMBH binaries are able to merge in both scenarios. In the empty loss cone case, we find that BHs merge via multi-body interactions, avoiding the 'final parsec' problem, and entering the PTA band with substantial orbital eccentricity. Our full loss cone treatment, albeit more approximate, suggests that the eccentricity becomes even higher when GWs become dominant, leading to rapid coalescences (binary lifetime $\lesssim 1$ Gyr). Despite the lower merger rates in the empty loss cone case, due to their higher mass ratios and lower redshifts, the GWB in the full/empty loss cone models are comparable (0.70×10^{-15} and 0.53×10^{-15} at a frequency of 1 yr^{-1}, respectively). Finally, we compute the effects of high eccentricities on the GWB spectrum.

5.1 Introduction

It is known that almost every nearby massive galaxy harbors a SMBH in its nucleus (Kormendy & Ho, 2013). In the ΛCDM cosmology, galaxies evolve as they hierarchically merge. As a result, it is expected that more than two SMBHs could coexist in a galaxy. If they successfully get close to each other, they form a bound pair. Recently, the presence of multiple SMBH systems has been observationally confirmed, such as a SMBH binary system at $z = 0.055$ with the projected separation of ~ 7 pc (Rodriguez et al., 2006; Bansal et al., 2017) and a triple SMBH at $z = 0.39$ with the closest pair separated by ~ 140 pc (Deane et al., 2014) .

However, it is still unknown whether the SMBH binary would further decay and eventually merge. This is one of the fundamental questions in astrophysics. The coalescence of two SMBHs, possibly being the loudest GW event in the Universe, has received much attention recently as we enter a new era of GW astronomy. In particular, Pulsar Timing Arrays (PTAs) are expected to be a powerful tool to detect the GWs emitted during the inpiral and

coalescence of SMBH binaries. Therefore, it is important to study formation and evolution of SMBH binaries.

Generally speaking, the evolution of a SMBH binary involves three stages from its formation to coalescence (Begelman et al., 1980). (i) As galaxies merge, SMBHs spiral to the core regions of the merged galaxies due to dynamical friction and form binaries. (ii) As the orbit shrinks, dynamical friction becomes inefficient and three-body interactions with surrounding stars or other orbiting BHs can cause the orbit of the SMBH binary to further decay. Viscous torques from a surrounding circumbinary disk can also play a role at this stage in a wet merger (e.g Mayer et al., 2007; Lodato et al., 2009; Mayer, 2013; Tang et al., 2017). (iii) Finally, at small enough separations, GW emission takes over, driving the SMBHs to merge. Whether or not SMBHs can merge is mostly determined by how smoothly and rapidly a transition from (i) to (iii) takes place[13].

In order for such transition to occur in less than the Hubble time, there must be a sufficient number of central stars to extract the orbital energy of the SMBH binary until it enters the GW-dominated regime. However, as the binary becomes more tightly bound, a significant fraction of stars are ejected, leaving behind empty phase space regions (the so-called empty "loss cone") around the binary with no stars remaining to interact with. The empty loss cone is replenished by dynamical processes, the simplest of which is two-body relaxation. Given the long relaxation time in the nuclei of bright elliptical galaxies ($\gtrsim$ 10 Gyr, Merritt & Wang 2005; Merritt et al. 2010), once the loss cone is cleared out, it is unlikely that it can be refilled fast enough - via two-body relaxation - to merge within a Hubble time. This may stall the SMBH binary at parsec scales, and is famously known as the "final parsec problem" (Milosavljević & Merritt, 2003). However, alternative dynamical mechanisms for sufficiently fast loss cone repopulation have been proposed, such as enhanced stellar flux into the core regions in non- spherical (triaxial or axisymmetric) nuclei (e.g Yu, 2002; Gualandris et al., 2017). Many studies have shown that some level of triaxiality is a characteristic of galatic merger remnants (e.g Preto et al., 2011; Khan et al., 2011, 2016; Gualandris & Merritt, 2012b) and that triaxial potential-density configurations can be dynamically stable over long timescales (e.g Poon & Merritt, 2002, 2004). This implies that aspherical geometries may prevent SMBH stalling, and yield a quick transition from phase (i) to (iii). In other words, SMBHs might coalesce on a shorter timescale than estimated assuming an empty loss cone.

However, it is also possible that the loss cone is not replenished efficiently. Because of resolution limitations, most N-body simulations of the final parsec problem in galaxy mergers are not likely converged (Vasiliev et al., 2014). More approximate Monte Carlo studies indicate that while axisymmetric potentials cannot solve the final parsec problem, realistic levels of triaxiality *can* (Vasiliev et al., 2015). However, triaxiality can erode over time due to chaotic diffusion (Merritt & Valluri, 1996), and, particularly in minor mergers, it is not clear that sufficient triaxiality is generated on small scales to refill the loss cone in time. Furthermore, the core/cusp dichotomy in the surface brightness profiles of galactic nuclei suggests that the (large) galaxies of greatest interest for pulsar timing efforts merge in a preferentially gas-poor way (Faber et al., 1997; Lauer et al., 2005). Interestingly, Dvorkin & Barausse (2017) suggest that in their extreme "nightmare scenario" in which SMBH binaries

[13]A bottleneck can arise earlier, in phase (i), for very small mass ratios $q_\star \ll 1$, when the dynamical friction time exceeds the Hubble time (Taffoni et al., 2003).

are assumed to stall and never complete their mergers, such a population of stalled binaries would produce a stochastic GW background at lower frequencies that should be detectable with PTAs .

The lack of theoretical consensus concerning solutions to the final parsec problem motivates us to consider the outcomes of stalled SMBH binaries *in a cosmological context*. If a SMBH binary fails to merge before another BH makes it to the nucleus as a result of a subsequent galaxy merger, multi-body interactions between the binary SMBH and the incoming BH will occur. Such triple BH interactions could be even more abundant at early times if more numerous SMBHs were assembled earlier, possibly promoting the formation of sufficiently compact binaries at high redshift which could merge by GW emission (Volonteri et al., 2003). The intrusion of another BH into the SMBH binary system can enhance the loss cone refilling rate by disturbing stellar orbits (Perets et al., 2007; Perets & Alexander, 2008). Moreover, chaotic, non-hierarchical three-body interactions tend to shrink the binary semimajor axis and to increase the eccentricity of an initially circular binary (Valtonen & Mikkola, 1991). If they form a hierarchical triple, the merger time of the inner binary can be dramatically reduced due to eccentricity oscillations induced by the Kozai-Lidov mechanism (Blaes et al., 2002). All of these effects likely accelerate the BH coalescence rate (Iwasawa et al., 2006; Bonetti et al., 2017) as well as the ejection rate of (typically less massive) SMBHs (Hoffman & Loeb, 2007). Ejection events - which can also occur due to GW recoil following successful SMBH mergers (Bekenstein, 1973; Campanelli et al., 2007a) - are observationally important for SMBH demographics (Schnittman, 2007; Kulkarni & Loeb, 2012). Understanding the outcomes of multiple SMBH interactions is therefore of great importance not just for determining merger rates, but also cosmological SMBH evolution.

In order to gain an in-depth understanding of SMBH binary evolution, observations of GWs using PTAs are crucial. 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). Their combined effort, the International PTA (IPTA, Hobbs et al. 2010), recently released its first datasets (Verbiest et al., 2016). With the duration of the observation $T \sim$ a few years to a few months and the observing cadence of $\Delta t \sim$ a few weeks, the relevant frequency band is beween $1/T$ and $1/2\Delta t$. This corresponds to approximately $n\mathrm{Hz} - \mu\mathrm{Hz}$. This frequency range is comparable to that of GWs from compact sub-parsec ($0.01 - 0.1$ pc) SMBH binaries. This makes the SMBHs one of the most promising astrophysical sources of GWs accessible to PTAs. A stochastic GW signal can be described by its amplitude $h_{\rm c}$, also known as the characteristic strain. In particular, for each individual SMBH binary in a circular orbit, it is easily shown that the strain scales as $h_{\rm c}(f) \propto f^{-2/3}$, where f is the observed frequency (Phinney, 2001). The strain is usually quoted at the frequency $f = 1\ \mathrm{yr}^{-1}$, and then referred to as A (Jenet et al. 2006; Eq. 103 of this paper). The stochastic GWB from massive BH mergers has been extensively examined via semi-analytical (e.g Wyithe & Loeb, 2003; Ravi et al., 2014) or Monte Carlo approaches (e.g Sesana et al., 2009; McWilliams et al., 2014; Kulier et al., 2015; Kelley et al., 2017), and it is typically estimated that $A \simeq (0.1 - 6) \times 10^{-15}$. However, so far, most of the studies have relied on the galaxy (or dark matter halo) merger history (merger rate and merger mass ratio) and assumed that the SMBH coalescence rates track the galaxy merger rates.

In this paper we adopt a *dynamical* approach to SMBH orbital evolution following merg-

ers, and we use it to estimate BH merger rates for both the full and the empty loss cone scenarios. Given the merger histories of galaxy samples in a mass range $M_\star = 10^{10} - 10^{12}$ $M_\odot$, for $0 < z < 4$, from the Millennium simulation (Springel, White, et al., 2005), we follow the evolution of SMBH binaries and their coalescences as the host galaxies go through minor/major mergers. Based on the inferred merger rates, we then predict the stochastic GW background. Our work is a first attempt to compute the global GWB by using few-body simulations to follow the dynamical evolution of multiple SMBH systems as a consequence of multiple galaxy mergers with a broad range of mass ratios ($10^{-4} < q_\star < 1$, where $q_\star$ is the mass ratio of two merging galaxies, defined to be smaller than 1). We explore two extreme scenarios for the last stage of the decay of a hard binary to bracket the range of outcomes to the final parsec problem: the full loss cone and the empty loss cone limits. In the empty loss cone case, dynamical friction no longer affects the evolution of the orbits when binaries become hard. We treat the full loss cone case in a more approximate way, assuming that dynamical friction always operates efficiently to cause orbital decay down to the merger. This is merely an approximation to the more complex physics of stellar scattering in the full loss cone regime (and furthermore neglects hydrodynamical solutions to the final parsec problem), but as we argue later, it is a reasonable approximation for high mass-ratio systems.

In our suites of simulations, we find that SMBH binaries merge in both scenarios, but with higher coalescence rates in the full loss cone case than in the empty loss cone one. In the full loss cone model, when GW-driven evolution becomes more dominant, the binary eccentricities are almost unity ($e > 0.99$), confirming some past predictions Quinlan (1996); Antonini & Merritt (2012). Subsequently, SMBH binaries coalesce rapidly (binary lifetimes $\lesssim 1$ Gyr). On the other hand, in the empty loss cone model, multi-body interactions of SMBHs play an important role in the decaying and coalescing of SMBH binaries. The binary lifetimes are longer ($\gtrsim 1$ Gyr). Using the inferred BH coalescence rates, we estimate A between the two models, $A = 0.70 \times 10^{-15}$ and $A = 0.53 \times 10^{-15}$ for the full loss cone and the empty loss cone case, respectively. They are comparable because (i) the higher coalescence rates of the full loss cone model come mostly from higher rates of low mass ratio mergers that contribute little to the GWB; high mass ratio systems merge in both models, (ii) more abundant and louder BH coalescence events at a later time (i.e. more massive mergers via mass growth and multi−BH interactions at small z), and (iii) the larger mass ratios of merged binaries, which increase the contributions of less massive binary mergers (in less massive galaxies) to the stochastic background signal, relative to the full loss- cone regime.

This paper is organized as follows. In §5.2, we explain our numerical setup including galaxy sampling (§5.2.1) and describe our model galaxies (§5.3) and prescriptions for BH mergers (§5.3.3 - 5.3.4). We present our results in §5.5. In §5.7, we estimate the stochastic GWB and further discuss the effects of high eccentricity on GW spectra. Finally, we conclude with a summary of our findings in §5.7.5.

5.2 Numerical Setup

In this section we describe the main ingredients of our galaxy/SMBH modelling. In particular, we detail how we select galaxy samples and how we treat galaxy mergers and the consequent rearrangement in the background potentials of DM and stars. We also describe

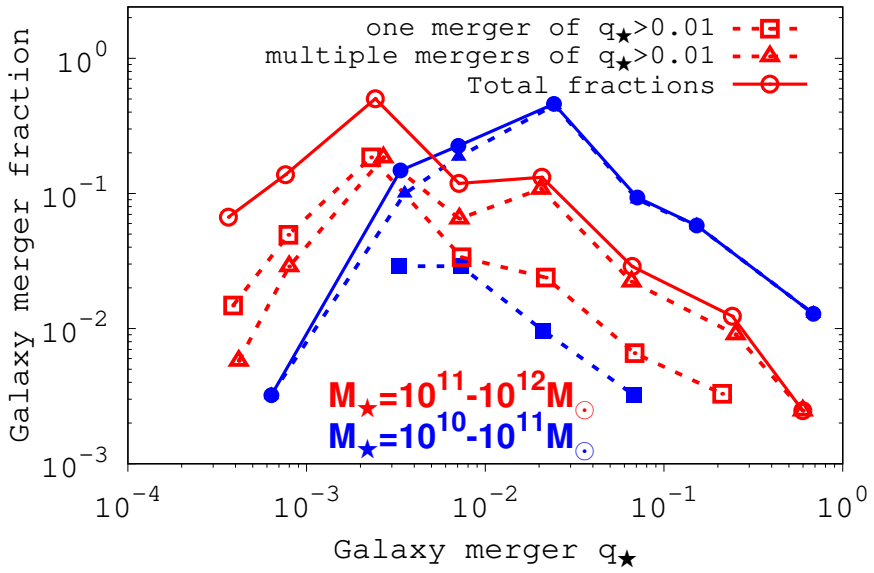

Figure 22: The fraction of galaxies that experience mergers (solid line with circles) with a given galaxy merger mass ratio $q_\star$. We distinguish the host galaxies by the number of significant mergers ($q_\star > 0.01$): the host galaxies with one significant merger (dotted line with squares) and those with multiple significant mergers (dotted line with triangles). Note that, while the lines for the galaxies experiencing no significant merger are not drawn, their contributions are included in the total fractions.

how we take into account the formation of SMBH binaries and how we define a BH merger.

5.2.1 Sampling of dark matter and galaxy merger trees

We follow merger trees of DM subhalos sampled from the Milli-Millennium simulation (Springel, White, et al., 2005)[14]. The Millennium simulation [15] is a large N-body simulation of cosmological structure formation performed with the GADGET-2 code assuming the standard ΛCDM cosmology with the cosmological parameters of $\sigma_m = 0.25, \sigma_b = 0.045, \sigma_\lambda = 0.75, h = 0.73$ and $\sigma_8 = 0.9$. The simulation follows the evolution of $N \approx 10^{10}$ particles in a periodic box of 500 h^{-1}Mpc on a side from $z = 127$ to $z = 0$. The simulation provides a total of 64 snapshots at redshifts from $z \approx 20$ to $z = 0$, equally spaced in $\log(1 + z)$. Throughout this paper, we assume that each DM halo hosts a galaxy whose mass is proportional to that

[14]http://gavo.mpa-garching.mpg.de/Millennium/

[15]There has been much progress in cosmological simulations since the Millennium: More advanced numerical techniques have been used in several simulations, such as 'Illustris' (Vogelsberger et al., 2014). Those simulations have successfully captured complicated effects induced by mutual interactions between gas, stars, DM and even BHs, which could not be achieved in DM-only simulations like the Millennium. However, the general physical picture should be shared by all those simulations, in particular the treatment of DM/galaxy mergers (e.g Rodriguez-Gomez et al., 2015), which is one of our main model ingredients. For the purpose of our study, the Millennium simulation allows us more freedom to choose/implement different model ingredients under the same physical framework of galaxy formation.

of the DM halo.

For galaxies at $z = 0$ (denoted by "host" galaxy) in the Millennium simulation, we follow the merger history of each host galaxy assuming a SMBH seed located at the center, from the past ($z > 0$) to the present day ($z = 0$). We will describe the detailed prescriptions for seed SMBHs in §5.3.3. The total number of the sampled host galaxies is 212. We consider galaxy mergers in each tree up to 10 - 12 for each host galaxy. This amounts to a total of 1733 galaxy mergers. The stellar masses of the host galaxies (the scaling relation 1 in §5.3.1) range within $M_\star = 10^{10} - 10^{12}$ $M_\odot$[16] (corresponding to virial masses of the host dark matter halos ranging from $M_{\rm DM,host} = 10^{12} - 10^{14}$ $M_\odot$). The earliest galaxy merger occurs at redshift $z = 3.58$ (or a cosmic time of $t = 1.76$ Gyr) and $z = 4.18$ (cosmic time of $t = 1.47$ Gyr) for host galaxies in the mass ranges of $M_\star = 10^{11} - 10^{12}$ $M_\odot$ and $M_\star = 10^{10} - 10^{11}$ $M_\odot$, respectively. In this paper, we refer to (smaller) galaxies merging with the host galaxies as "satellite" galaxies of mass $M_{\star,\rm sat}$. The galaxy merger ratio $q_\star$ between the satellite and host galaxy is defined to be smaller than 1, namely $q_\star = M_{\star,\rm sat}/M_{\star,\rm host}$.

In Figure 22, we show the fraction of galaxies that go through mergers (solid line with circles) with a given $q_\star$. We sub-categorize the host galaxies into two bins depending on the number of significant mergers ($q_\star > 0.01$) [17] they experience: the host galaxies with one significant merger and those with multiple significant mergers. We find that mergers with $q_\star < 0.01$ are more common for galaxies of $M_\star = 10^{11} - 10^{12}$ $M_\odot$ (the maximum count at $q_\star \sim 3 \times 10^{-3}$). For galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$, the most frequent are mergers of $q_\star \sim 3 \times 10^{-2}$ and the $q_\star$ distributions have shorter low-$q_\star$ tails. In addition, we can see that the majority of the host galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ experience more than two major mergers. Therefore, all of these indicate that the merger mass ratio $q_\star$ is generally higher for less massive host galaxies. This may imply that SMBHs merge with relatively high rates in galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ compared to more massive galaxies, which will be shown in §5.6. We also summarize those merger counts in Table 11.

[16] In this paper, the subscript $\star$ indicates physical quantities related to galaxies, while quantities with a subscript "BH" or without one refer to the SMBHs. For example, we have galaxy masses $M_\star$, but BH masses $M_{\rm BH}$. Similarly, galaxy merger mass ratios are indicated with $q_\star$, while BH binary mass ratios with q.

[17] Throughout this paper, we only use the terms "significant ($q_\star > 0.01$)", "major ($q_\star > 0.25$)" and "minor ($q_\star < 0.25$)" mergers to refer to galaxy mergers.

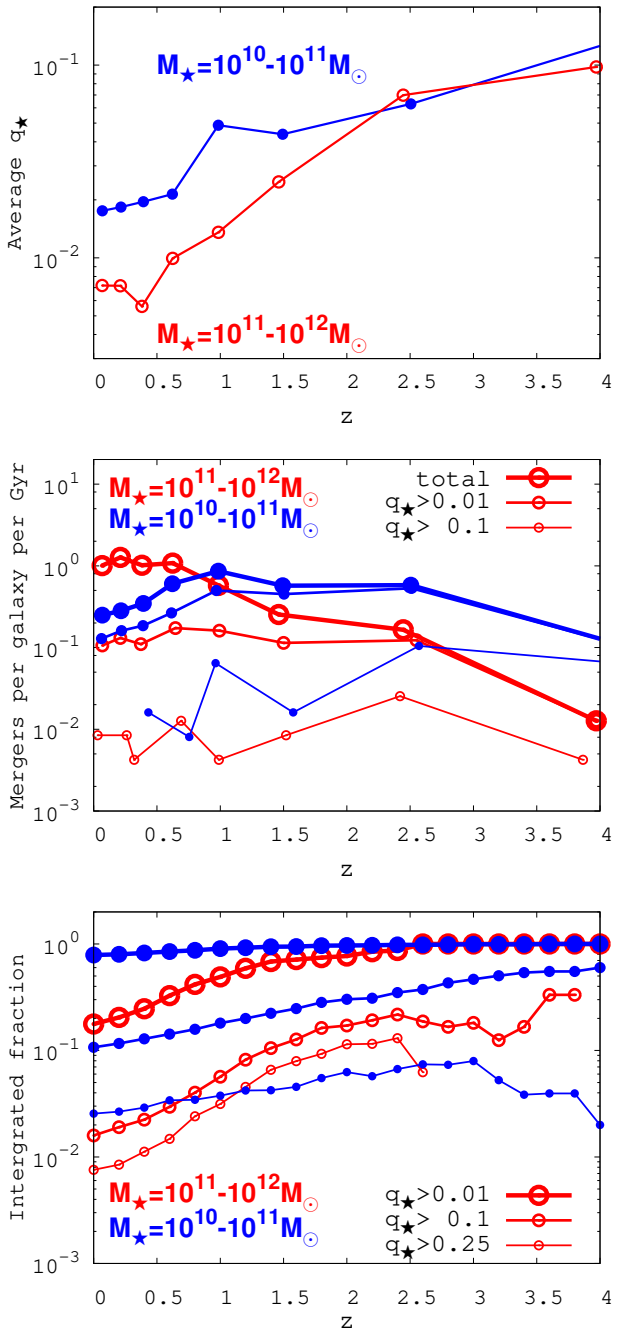

Figure 23: The evolution of the galaxy merger mass ratios (or the stellar mass ratio) $q_\star$, galaxy mergers and their number fractions as a function of redshift. The *upper* panel shows the merger mass ratios averaged per Gyr as a function of redshift. Also, we present in the *middle* panel the number of mergers per galaxy per Gyr. The line thickness indicates different, progressive cutoffs on the mass ratio: from the merger count without any cutoff (thickest line) to the mergers of $q_\star > 0.1$ (thinnest line). In the *bottom* panel, we present the number of significant galaxy mergers normalized by the total galaxy merger counts up to a given redshift.

Table 11: Overview of our galaxy samples. From top to bottom: The mass of sampled galaxies, the total number of host galaxies, the total number of galaxy mergers and BHs. In the last four rows, we present the number fraction of host galaxies experiencing mergers with different mass ratios $q_\star$ and frequencies $N_\star$: (from the $4^{\rm th}$ to the last row) one significant merger of $q_\star > 0.01$, more than one significant mergers of $q_\star > 0.01$, at least one major merger of $q_\star > 0.25$ and $q_\star > 0.5$.

Galaxy mass $M_\star$		$10^{10} - 10^{11}$ M$_\odot$	$10^{11} - 10^{12}$ M$_\odot$
Total number of host galaxies (at $z = 0$)		75	137
Total number of galaxy mergers / BHs		462 / 457	1271 / 1256
Number fraction of host galaxies with one significant merger	$(N_\star = 1, q_\star \geq 0.01)$	7%	32%
Number fraction of host galaxies with more than one significant mergers	$(N_\star \geq 2, q_\star \geq 0.01)$	93%	42%
Number fraction of host galaxies with at least one major merger	$(N_\star \geq 1, q_\star \geq 0.25)$	12%	7%
Number fraction of host galaxies with at least one major merger	$(N_\star \geq 1, q_\star \geq 0.5)$	8%	2%

We present in Figure 23 the average mass ratio $q_\star$, the merger rate per galaxy and the merger fraction, as a function of redshift. The *upper* panel shows the merger mass ratios averaged per Gyr as a function of redshift. It can be seen that the merger mass ratio is generally higher for galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$, independent of redshift. In the *middle* panel we show the number of mergers per galaxy per Gyr. The thickness of the line represents progressive, different cutoffs on the mass ratio: from the merger count without any cutoff (thickest line) to the mergers of $q_\star > 0.1$ (thinnest line). In the *bottom* panel we present the cumulative distribution of significant mergers in z, or the number fraction of significant mergers integrated up to a given redshift.

5.3 Model description

In this section we describe our modelling of DM and galaxy potentials, as well as the treatment of SMBHs, and in particular their seed masses, their orbital parameters at galaxy mergers, and their mass growth. Furthermore, we describe our treatment of dynamical friction and the BH merger conditions using two different prescriptions.

5.3.1 Dark matter and stellar distribution and seed BH mass

We model gas-poor galaxies with three components: DM, stars and SMBHs. As we follow the merger histories of the host galaxies in the Millennium simulation, at every galaxy merger DM and stellar potentials are re-established based on the new mass of the galaxy after the merger.

We adopt the NFW profile for the DM density distribution $\rho_{\rm DM}$ with concentration parameter $C = 3$ (e.g Van Wassenhove et al., 2014). For numerical convenience, we slightly modify the inner region of the NFW profile ($\rho_{\rm DM} \sim r^{-1}$) so that the DM density does not exceed the stellar density at the very center of the galaxy core. This only affects the region inside $\sim (10^{-3} - 10^{-4}) r_{\rm c}$, and does not appreciably affect our results.

We consider the stellar density distribution for merged galaxies explored in Stone & Ostriker (2015):

$$\rho_\star = \frac{\rho_{\rm c}}{(1 + r^2/r_{\rm c}^2)(1 + r^2/r_{\rm h}^2)}\,, \tag{85}$$

where $\rho_{\rm c}$ is the central density, $r_{\rm c}$ is the core radius and $r_{\rm h}$ is the outer halo radius (or half-mass radius). The profile has a flat central core in the innermost region ($r < r_{\rm c}$), smoothly extending outward with $\rho_\star \propto r^{-2}$ for $r_{\rm c} \leq r < r_{\rm h}$ and $\rho_\star \propto r^{-4}$ for $r_{\rm h} \leq r$.

The post-merger stellar density profile $\rho_\star(r)$ of a merged galaxy is more complex than this idealized model, but our choice of $\rho_\star(r)$ is motivated by observations of large elliptical galaxies that are likely the primary hosts of PTA sources. Specifically, *Hubble Space Telescope* (*HST*) observations of the nuclear regions of nearby early type galaxies find a bimodality in surface brightness profiles $I(R)$ (here R is a projected 2D radius, as opposed to the 3D radial coordinate r). When power law profiles are fit to the inner isophotes of *HST* data, i.e. $I(R) \propto R^{-\Gamma}$, the resulting Γ distribution is strongly bimodal, with most galaxies having either $0 < \Gamma < 0.3$ or $0.5 < \Gamma < 0.9$ (Lauer et al., 2005). The former type of galactic nucleus, known as a "core" profile, is dominant among galaxies brighter than $M_V \approx -20$ (Graham et

al., 2003; Graham & Guzmán, 2003), and is roughly consistent with the flat inner slope one obtains by projecting Equation 85.

Flat cores in surface brightness profiles could be created by the dynamical effects of SMBH binaries. In the aftermath of a galaxy merger, hosted SMBHs are effectively dragged towards the centre of the merged galaxies by dynamical friction, and eventually form a binary. The binary acts as a heating source as its orbit shrinks, pumping the lost energy to the background stellar populations. The deposition of the binary's orbital energy can scour out a flat core of stars in the inner region, creating a mass deficit relative to the initially steeper density profile (e.g. see Chapter 7 in Merritt, 2013, and references therein). The creation of flat cores by SMBH binaries has been confirmed in numerical simulations (e.g. Merritt, 2006; Gualandris & Merritt, 2012b; Kulkarni & Loeb, 2012; Bortolas et al., 2016). Furthermore, stellar scouring has been inferred in a number of core elliptical galaxies from observations (e.g. Thomas et al., 2014), and is widely predicted in numerical studies (e.g. Milosavljević & Merritt, 2001; Kormendy & Ho, 2013). Although both dynamical friction and three-body stellar scatterings contribute to core creation in the vicinity of an SMBH binary, we only include the former (Ebisuzaki et al., 1991) in our model[18]. We discuss limitations of our simple treatment of phase (ii) later in this section.

For a given DM halo mass $M_{\rm DM,host}$ at redshift z, the DM density distribution is completely determined. However, we have three free parameters for the stellar potential to be fixed, namely, $r_{\rm c}$, $r_{\rm h}$ and $\rho_{\rm c}$. In order to fix those parameters as well as the seed SMBH mass, we solely depend on four observational scaling relations:

1. $M_{\rm DM} - M_\star$ relation : $\left(\frac{M_{\rm DM}}{10^{13}\ {\rm M}_\odot}\right) = 0.50 \left(\frac{M_\star}{10^{11}\ {\rm M}_\odot}\right)$
   (Lin et al., 2012; Kulier et al., 2015)

2. $M_{\rm BH} - M_\star$ relation : $\left(\frac{M_{\rm BH}}{10^{9}\ {\rm M}_\odot}\right) = 0.49 \left(\frac{M_\star}{10^{11}\ {\rm M}_\odot}\right)^{1.16}$ (Kormendy & Ho, 2013)

3. $M_{\rm BH} - \sigma$ relation : $\left(\frac{M_{\rm BH}}{10^{9}\ {\rm M}_\odot}\right) = 0.309 \left(\frac{\sigma}{200\ {\rm km\ s^{-1}}}\right)^{4.38}$ (Kormendy & Ho, 2013)

4. $M_{\rm BH} - r_{\rm c}$ relation : $\left(\frac{r_{\rm c}}{\rm kpc}\right) = 0.0821 \left(\frac{M_{\rm BH}}{10^{9}\ {\rm M}_\odot}\right)^{0.855}$ (Thomas et al., 2016),

We note that we ignore the scatter in the above relations, assuming them to be exact. [19] While the general $M_{\rm DM} - M_\star$ relation is more complicated than our prescription (e.g Moster et al., 2013), a single power law is a reasonable approximation to the high mass end of this relation that we focus on. With these relations, the DM, the stellar distributions, and the seed SMBH mass are determined given the DM halo mass. In particular, we assume that central SMBHs are missing in galaxies of $M_\star < 10^8\ {\rm M}_\odot$, which correspond to the minimum

[18] The anisotropic emission of GWs (or "gravitational rocket effect") during the final coalescence of two SMBHs may also produce a mass deficit in galactic nuclei (Merritt et al., 2004; Gualandris & Merritt, 2008) following recoil of the merged SMBH, but we neglect this in our model.

[19] Taking into account scatter and the impact of different choices of the SMBH-galaxy relation to populate SMBHs may overly introduce complexities to our analysis. For simplicity, we neglect scatter in galaxy scaling relations used in our model. However, this issue has been addressed before, for example in (Shankar et al., 2016; Sesana et al., 2016; Rasskazov & Merritt, 2017).

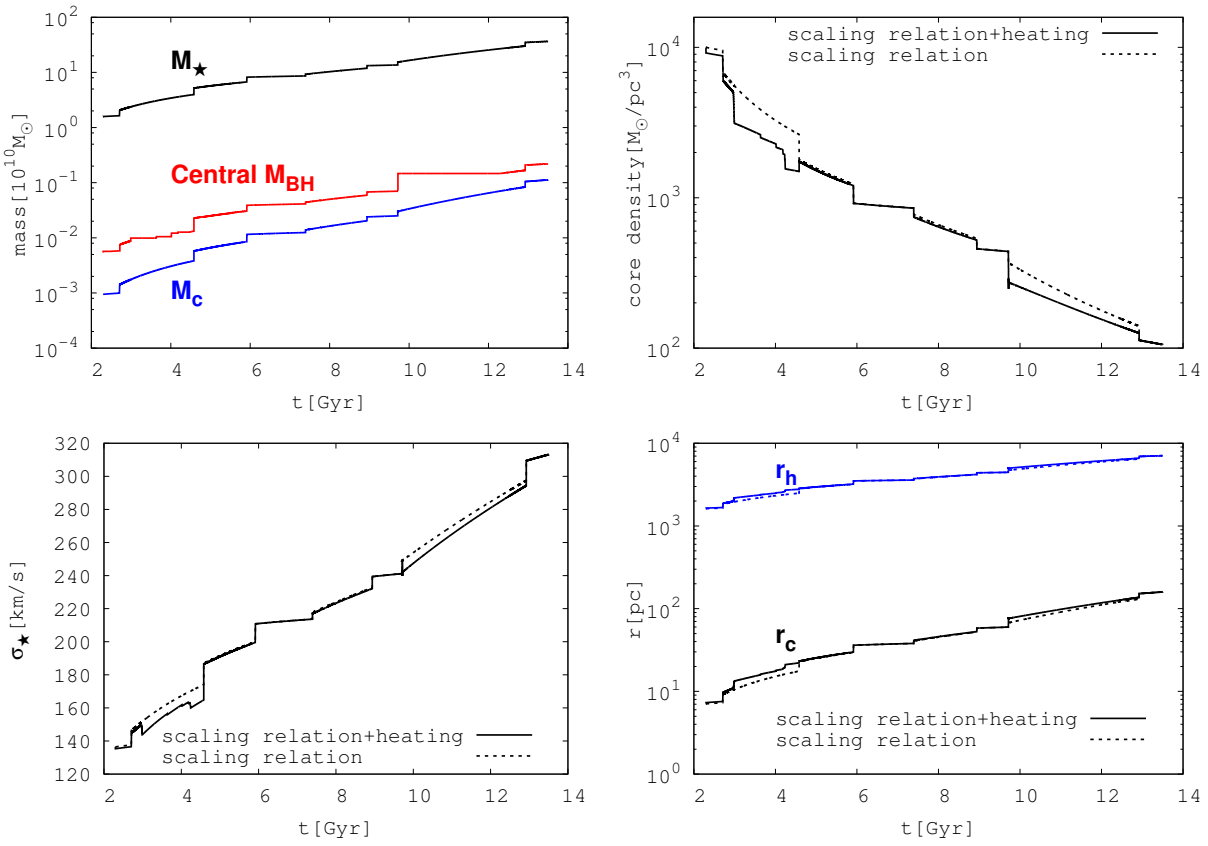

Figure 24: The evolution of stellar mass ($M_\star$), central BH mass (central $M_{\rm BH}$) and core mass ($M_{\rm c}$) (*top-left*), core density $\rho_{\rm c}$ (*top-right*), $\sigma_\star$ (*bottom-left*) and two characteristic radii $r_{\rm h}$ and $r_{\rm c}$ (*bottom-right*) of one massive galaxy among the sampled galaxies. We show those variables as determined only by the scaling relations (dotted lines) as well as when the heating effect due to dynamical friction is additionally taken into account (solid lines). The host galaxy grows via 9 mergers from $M_\star \simeq 10^{10}$ M$_\odot$ at $z = 2.8$ ($t = 2$ Gyr) to $M_\star \simeq 3\times10^{11}$ M$_\odot$ at $z = 0$ ($t = 13.8$ Gyr). The mass of the central (most massive) BH has reached $M_{\rm BH} \simeq 10^9$ M$_\odot$ at $z = 0$. Overall, the core swells ($r_{\rm c}$ and $M_{\rm c}$) as $M_\star$ increases, but ρ_c declines.

BH mass $M_{\rm BH} = 10^{5.5}$ M$_\odot$ in our simulations. For mergers with such small galaxies, we simply add the masses of the small galaxies to the host galaxy masses without placing the seed SMBHs. Those small galaxies occupy around 1% of the total number of satellite galaxies for all galaxy mass ranges.

In order to see how the stellar potential evolves as the total stellar mass increases, we express $r_{\rm c}$, the core stellar mass $M_{\rm c}$ and $\rho_{\rm c}$ in terms of $M_\star$:

$$r_{\rm c} \propto M_\star^{0.99}\,, \tag{86}$$

$$M_{\rm c} \propto M_\star^{1.52}\,, \tag{87}$$

$$\rho_{\rm c} \propto M_\star^{-1.46}\,. \tag{88}$$

The relations are derived in Appendix B and imply that as galaxies (DM subhalos) grow in mass, the core regions expands in size and mass whereas the core density declines (Dullo & Graham, 2014). Even though the dependencies on $M_\star$ differ, those trends are consistent with those of Faber et al. (1997), i.e. $r_{\rm c} \sim M_\star^{0.92}$, $M_{\rm c} \sim M_\star^{1.24}$ and $\rho_{\rm c} \sim M_\star^{-1.52}$. As an example, in Figure 24 we show the evolution of $M_\star$, the central BH mass and $M_{\rm c}$, $\rho_{\rm c}$, $\sigma_\star$ and $r_{\rm h}$ and $r_{\rm c}$ of one of the more massive galaxies in our sample. In the plots, we show those variables as determined only by the scaling relations as well as when the heating effect due to dynamical friction (see below) is additionally taken into account. The host galaxy grows via 9 mergers from $M_\star \simeq 10^{10}$ M$_\odot$ at $z = 2.8$ ($t = 2$ Gyr) to $M_\star \simeq 3 \times 10^{11}$ M$_\odot$ at $z = 0$ ($t = 13.8$ Gyr). The mass of the central (most massive) BH has reached $M_{\rm BH} \simeq 10^9$ M$_\odot$ at $z = 0$. As expected, the core swells ($r_{\rm c}$ and $M_{\rm c}$) as $M_\star$ increases, but ρ_c declines. Notice that $r_{\rm h}$ has a relatively weak dependence on $M_{\rm c}$ compared to $r_{\rm c}$, i.e., $r_{\rm h} \sim M_\star^{0.47}$. We go into greater detail on evolutionary deviations due to heating effects and the prescription for BH mass growth in the following section (§ 5.3.2).

We provide in Table 13 in Appendix B the scaling relations between the relevant variables in our model in terms of $M_{\rm BH}$ (as well as $M_{\rm DM}$), derived with the four scaling relations 1-4.

5.3.2 Evolution of DM and galaxy potential and BH mass growth

As galaxies merge, DM and stellar potentials evolve in time. For the DM potential, we interpolate the DM halo masses between two adjacent galaxy mergers (or two different redshifts or snapshots at mergers) in the Millennium simulation. In particular, we use a fitting formula derived by Wechsler et al. (2002), which can be written as follows,

$$M_{\rm DM}(z) = M_{\rm DM,0} \exp\left[-\Delta\left(\frac{z+1}{z_0+1} - 1\right)\right]\,, \tag{89}$$

where z_0 is the redshift when a halo is observed. Here, we assume z_0 to be the same as the redshift at which two halos merge in the Millennium simulation. Therefore, given the mass of a merged halo (or merged galaxy) at $z = z_0$ and the subsequent merger at $z = z_1$, we determine Δ[20]. In the Millennium simulation, DM halos typically grow in mass from one

[20]More explicitly, Δ is expressed in Wechsler et al. (2002) as $S/(1 + z_{\rm c})$, where $z_{\rm c}$ is the redshift at which the halo collapses and S is a characteristic factor which relates the accretion rates of halos.

merger to the following merger. However, there are also cases where the DM halo masses at subsequent mergers are found to be smaller. On average, sampled host galaxies experience such decreases in mass once in their merger histories. This could be caused by several mechanisms, and in particular tidal stripping. But the precise cause cannot be determined from the information provided in the snapshots alone. In this paper, for such cases, we conservatively assume that DM halo masses do not change between the two mergers, but we update the halo masses accordingly at the later merger. In addition to the growth of DM halos, we also take into account the widening of the stellar potential due to the "scouring effect" (Milosavljević et al., 2002; Merritt, 2006) of SMBH binaries as a result of dynamical friction. As the orbits of SMBH binaries shrink, they lose their energy to background stars, which will clear out some stars onto wider orbits. To quantify this effect on the stellar potential [21] we compute, at every time step Δt, the dissipative energy $E_{\rm dis,i}$ due to the dynamical friction force $\boldsymbol{f}_{\rm df,i}$ (see Equation 97) for the $i_{\rm th}$ BH moving at velocity $\boldsymbol{v}_i$,

$$E_{\rm dis} = \sum_i |\boldsymbol{f}_{{\rm df},i} \cdot \boldsymbol{v}_i| \Delta t. \tag{90}$$

Hence we deposit $E_{\rm dis}$ into the virialized stellar potential assuming the total mass of stars $M_\star$ is fixed and the three-parameter structure of the density distribution is maintained. By the virial theorem, the total potential energy of stars $W_\star$ can be expressed in terms of the total binding energy of stars $E_\star$, the dissipative energy $E_{\rm dis}$ and the virially averaged dispersion $\sigma_\star$ as follows,

$$-W_\star = -2(E_\star + E_{\rm dis}) = M_\star \sigma_\star^2. \tag{91}$$

Stone & Ostriker (2015) provide the explicit expressions for the total potential energy $W_\star$ (Equation 8) and $M_\star$ (Equation 5) in terms of $\rho_{\rm c}$, $r_{\rm c}$, $r_{\rm h}$ and $\sigma_\star$. With the scaling relations 1-4, we can then estimate the adjusted values of $\rho_{\rm c}$, $r_{\rm c}$ and $r_{\rm h}$, and update them accordingly at every time step. Given $W_\star$, $E_\star < 0$, the scouring effect produces an expansion of the characteristic size of the potential ($r_{\rm c}$ and $r_{\rm h}$), while lowering the core density $\rho_{\rm c}$, as shown in Figure 24. However, note that the decrease in the core density is accompanied by mass growth of the galaxies.

In our simulations, the masses of the central BHs increase such that $M_{\rm BH} - \sigma$ (scaling relation iii) is always satisfied. The central BHs are defined in this paper as BHs whose entire orbits (either with respect to galaxy potential or in binaries with other BHs) are confined to the core. If BHs only temporarily stay in the core region at their closest approach (pericentre) to the origin, they are not identified as central BHs. In our simulations, we find that the central BHs typically include the most massive BHs (denoted by $\rm BH_1$ and their masses $M_{\rm BH,1}$) and the BHs forming bound pairs with $\rm BH_1$. The total mass of the central BHs (denoted by $M_{\rm cBH}$) is mostly dominated by $M_{\rm BH,1}$. If the BH mass required by $M_{\rm BH} - \sigma$ (denoted by $M_{{\rm BH},\sigma_\star}$ where $\sigma_\star$ is the virially averaged dispersion defined in Equation 91 [22])

[21] Our treatment of scouring only alters the stellar, not the DM, density profile. However, our results are not significantly affected by whether or not the DM density profile is influenced by scouring effects since the stellar potential is dominant near the core regions where binaries and multiple BHs interact.

[22] Note that the variable σ used in the $M_{\rm BH} - \sigma$ relation may not have exactly the same meaning as $\sigma_\star$

is already smaller than $M_{\rm cBH}$, the mass of each central BH stays the same. On the other hand, for $M_{\rm BH,\sigma_\star} > M_{\rm cBH}$, given the mass $M_{\rm BH,i}$ of each central BH$_i$ at a certain time step, the mass of the BH$_i$ at the following time step $M'_{\rm BH,i}$ increases by a factor of $M_{\rm BH,\sigma_\star}/M_{\rm cBH}$, or simply,

$$M'_{\rm BH,i} = M_{\rm BH,i}\frac{M_{\rm BH,\sigma_\star}}{M_{\rm cBH}}. \tag{92}$$

With this crude approximation for mass growth through gas accretion we ensure that the masses of the central BHs are maintained at realistic values. On the other hand, whenever the central massive BHs are missing in the core regions, given the mass reservoir in these regions, the masses of other small BHs, which fall into the core later or have already existed, could grow rapidly up to masses comparable to the missing central BHs. In particular, in our models assuming instantaneous formations of post-merger galaxies, this can take place while the central BHs are dislocated off center at galaxy mergers. Only two such cases occurred in our simulation suite, and to be conservative, we exclude the contribution of these to the GWB (see §5.7).

5.3.3 Initial orbital parameters of SMBHs at galaxy mergers

In hierarchical models of structure formation in cosmology, DM halos grow via mergers as well as accretion of DM. During the process of merging, the orbital properties of infalling satellite halos have been investigated in many studies. Recent cosmological N-body simulations show that two halos typically merge on almost parabolic orbits with large eccentricity for various ranges of the halo mass, mass ratio and redshift (Benson, 2005; Khochfar & Burkert, 2006; Wetzel, 2011; Jiang et al., 2015). Additionally, studies of SMBH binary formation in merging galaxies generally assume such radial orbits for infalling SMBHs as initial conditions (Kulkarni & Loeb, 2012; Van Wassenhove et al., 2014; Capelo et al., 2015). Motivated by those studies, we also assume the initial orbits of incoming BHs with respect to the merged galaxy potential to be highly eccentric. In particular, we adopt a fitting formula for the eccentricity given in Wetzel (2011). Using cosmological N-body simulations, Wetzel (2011) investigated the orbital parameters of infalling satellite halos and their dependences on the halo mass and redshift. The author provides a simple functional form of the orbital distribution of the satellite circularity η for $z = 0-5$ and $M_{\rm DM,host} = (10^{10}-10^{15})h^{-1}$ M$_\odot$. The distribution of the circularity $df/d\eta$ adopted in this study is expressed as follows,

$$\frac{df}{d\eta} = 3.38\left(1+0.567\left[\frac{M_{\rm DM,host}}{M_0}\right]^{0.152}\right) \times \eta^{1.05}(1-\eta)^{0.242\left(1+2.36\left[\frac{M_{\rm DM,host}}{M_0}\right]^{0.108}\right)}, \tag{93}$$

of the virially averaged dispersion. However, considering systematic uncertainties in the dispersion measure (Tremaine et al., 2002), we conservatively assume the virially averaged value of $\sigma_\star$ as a representative for the dispersion of the host galaxy. Stone & Ostriker (2015) show that the virially averaged dispersion is comparable to the central dispersion for $r_{\rm h} \gg r_{\rm c}$.

where $\eta = \sqrt{1-e^2}$ and $\log[M_0/h^{-1}\ \mathrm{M}_\odot] = 12.42 - 1.56z + 0.038z^2$. We estimate the eccentricity using $e = (r_\mathrm{a} - r_\mathrm{p})/(r_\mathrm{a} + r_\mathrm{p})$, where r_a and r_p are the apocentric and pericentric distances of BH orbits, respectively, with respect to the galactic potential. For simplicity, the initial eccentricity is given in the simulations as the peak value of the fitting formula, $e \simeq 0.8 - 0.9$ at mergers.

For a merger between a host galaxy already hosting several BHs (BH_i with $i \geq 1$) and an incoming j_th satellite galaxy, we assume only one BH per satellite galaxy but we allow multiple mergers at the same redshift (i.e., $j \geq 1$). In the Millennium simulation, when more than two galaxies disappear from one snapshot to the next one, we assume that they merge with the host galaxy at the same time. For galaxy mergers at a given redshift, we find a post-merger galaxy system of a pre-existing BH cluster in a host galaxy and incoming BHs in satellite galaxies; these are found at the apocenter of their instantaneous orbits in the new spherical post-merger potential, re-established around the center of mass (CoM) of all BHs. The CoM of the pre-existing BH cluster and each of the incoming BHs are separated by $r \sim r_\mathrm{h}$ [23]. For the given initial positions (i.e., the apocenters of the initial orbits), the initial velocities are assigned to give highly eccentric orbits as above. Finally, the positions and velocities of BHs in host galaxies ($\mathbf{x}^{\mathrm{BH}'}_{\mathrm{host},i}$,$\mathbf{v}^{\mathrm{BH}'}_{\mathrm{host},i}$) and in the j_th satellite galaxy ($\mathbf{x}^{\mathrm{BH}'}_{\mathrm{sat},j}$,$\mathbf{v}^{\mathrm{BH}'}_{\mathrm{sat},j}$) are expressed for any number of mergers ($j \geq 1$) at a given redshift as follows,

$$
\begin{aligned}
\mathbf{x}^{\mathrm{BH}'}_{\mathrm{host},i} &= \mathbf{x}^{\mathrm{BH}}_{\mathrm{host},i} + \frac{\sum_j M_{\star,j}}{M_{\star,\mathrm{host}} + \sum_j M_{\star,j}} \times r_\mathrm{h} \hat{\mathbf{x}}^{\mathrm{BH}}_{\mathrm{host},i} \\
\mathbf{x}^{\mathrm{BH}'}_{\mathrm{sat,j}} &= \frac{M_{\star,\mathrm{host}}}{M_{\star,\mathrm{host}} + \sum_j M_{\star,j}} \times r_\mathrm{h} \hat{\mathbf{x}}^{\mathrm{BH}}_{\mathrm{sat},j} \\
\mathbf{v}^{\mathrm{BH}'}_{\mathrm{host,i}} &= \mathbf{v}^{\mathrm{BH}}_{\mathrm{host},i} \times \xi(q,n) + \sqrt{\frac{GM_\mathrm{en}(r < x^{\mathrm{BH}'}_{\mathrm{host,i}})}{x^{\mathrm{BH}'}_{\mathrm{host,i}}}(1-e)\alpha} \times \hat{\mathbf{v}}^{\mathrm{BH}}_{\mathrm{host},i} \\
\mathbf{v}^{\mathrm{BH}'}_{\mathrm{sat,j}} &= \sqrt{\frac{GM_\mathrm{en}(r < x^{\mathrm{BH}'}_{\mathrm{sat},j})}{x^{\mathrm{BH}'}_{\mathrm{sat},j}}(1-e)\alpha} \times \hat{\mathbf{v}}^{\mathrm{BH}}_{\mathrm{sat},j}\,,
\end{aligned}
$$

where $\boldsymbol{x}_i$ and $\boldsymbol{v}_i$ (without prime symbol) are the position and velocity vectors of BH_i just before mergers, and $\hat{\mathbf{x}}_i$ and $\hat{\mathbf{v}}_i$ are the randomly-generated unit vectors, satisfying $\hat{\mathbf{x}}_i \perp \hat{\mathbf{v}}_i$ (same for j as well). $M_\mathrm{en}(r < x')$ is the enclosed mass inside of $r = x'$ and α is a factor used to assign the eccentricity for the first orbit in a non-Keplerian potential (See equation 85). We conservatively use $\alpha \simeq 1/5$ for the eccentricity ranges given by Equation 93, i.e., $e > 0.8$ [24]. Here, we introduce a function $\xi(q,n)$ to quantify the extent by which a host galaxy is disrupted by a merger. We define the function $\xi(q,n)$ as a degree of memory for

[23] Note that for multiple mergers ($j \geq 2$), the separation between the two BH systems is not exactly r_h since they are not aligned on a line, but rather spread on a 2−dimensional plane ($j = 2$) or in a 3−dimensional space ($j \geq 3$).

[24] For the same velocity (not the circular velocity) at the same apocenter, the first pericenter distances are different in the Keplerian and Non-Keplerian potentials (i.e., different eccentricities). Therefore, some extra factor should be taken into account in the expression for v at apocenter in the Keperian potential. The value of α taken in this paper is comparable to that for the logarithmic potential ($\rho \sim r^{-2}$) (Innanen et al., 1982). Recall that our stellar density approximately follows $\rho \sim r^{-2}$ at $r < r_\mathrm{h}$.

the orbits of existing BHs in the host galaxies at given mergers, scaling from 0 (complete loss of memory) to 1 (complete retention of memory). Motivated by the considerations below, we define $\xi(q, n)$ assuming the following functional form,

$$\xi(q, n) \equiv \left| \frac{q^n - 1}{q^n + 1} \right| , \tag{94}$$

where $q = \sum_j M_{\star,j}/M_{\star,\rm host}$. During the process of merger, it is more likely that the system of host galaxies is disrupted by mergers of high q. In other words, as they go through major mergers, the host galaxies lose memory of the dynamics before the mergers ($\xi \simeq 0$ for $q \to 1$). BHs in the host galaxies, however, are less influenced by minor mergers, possibly keeping more memory of the dynamics ($\xi \simeq 1$ for $q \to 0$). n is meant to inform how much the dynamics of BHs in the host halo is affected by a given galaxy merger. For this study, we conservatively take $n = 1$. We hope that a more precise functional form will be found in future studies.

We note that with the prescriptions for $\boldsymbol{v}$ and the assumption of instantaneous formation of post-merger galaxies, the orbits of pre-existing BHs become possibly either more radial or more circularized at mergers. We further note that it is possible that BHs could escape from the potential or their apocenters could become significantly larger than $r_{\rm h}$ if they happened to gain sufficient kinetic energies at mergers. However, in our simulations, we could not find such cases.

5.3.4 BH merger conditions

The fate of the SMBHs after galaxy mergers is still not fully understood, with uncertainties remaining on whether SMBH mergers do occur, and on which timescale. However, under the assumption that SMBHs do eventually merge, it is important to estimate how frequently they do so given the merger histories of the host galaxies. At large separations, dynamical friction plays a dominant role in bringing two massive BHs together to form a bound binary. As they become more tightly bound, a significant amount of stars may be ejected, leaving behind an empty loss cone. Given the long relaxation time in the nuclei of early-type galaxies ($\sim$ 10 Gyr, Merritt 2006), once the stars are cleared out, it is unlikely that collisional processes can refill the loss cone before $z = 0$. Many alternative mechanisms to solve the final parsec problem exist, from nuclear triaxiality to circumbinary disks (see §5.1). Treating all of these mechanisms in a self-consistent way is far beyond the scope of this paper, which primarily aims at studying the role of multi-SMBH interactions in the solution of the final parsec problem. We therefore focus only on dynamical friction and multi-SMBH encounters as drivers of orbital evolution.

We consider two extreme scenarios for dynamical friction. In our fiducial model, we assume dynamical friction stops affecting SMBH orbits once binaries become sufficiently tight. We refer to this as the "empty loss cone model", or "ELC-model" for short. In the ELC-model, if binaries satisfy any of the following conditions, dynamical friction is deactivated:

1. Hard binary: when the semimajor axis of the BH binary is smaller than the hard semimajor axis $a_{\rm h}$, or $a < a_{\rm h} = G\mu/4\sigma_\star^2$ (μ is the reduced mass of the binary);

2. Fast-moving stars: when the speeds of the BHs are slower than the local circular velocity, or $v < \sqrt{G[M_{\rm en}(r) + M_{\rm BH}(r)]/r}$;

3. Inside the influence radius $r_{\rm in} = 2GM_{\rm BH,1}/\sigma_\star^2$ (where $M_{\rm BH,1}$ is the primary BH mass): when a less massive BH in a binary is inside the influence radius of a more massive BH but no 3$^{\rm rd}$ BH is inside $r_{\rm in}$.

The ELC-model is meant to investigate multi-SMBH interactions as a "mechanism of last resort" for solving the final parsec problem in massive galaxies where alternative solutions are likely to be less reliable.

In our alternative scenario, we assume that dynamical friction always play a role until binaries merge. We refer to this case as "full loss cone model" or simply "FLC-model" [25]. We emphasize that our FLC-model assumes full loss cones and the standard Chandrasekhar formula (see Equation 97) as a valid way to evaluate dynamical friction for hard binaries in the full loss cone regime. The standard Chandrasekhar formula was derived under the assumption of non-accelerated/linear motion in a uniform density distribution. When a binary enters the hard-binary regime, as the gravity from the second binary becomes more important, those assumptions of the dynamical friction formula may not be valid any more. However, by continuing to use the usual dynamical friction formula in the FLC-model down to the GW-driven regime, we ignore these corrections. We discuss the analytic validity, as well as the limits and caveats of full loss cone assumption in more detail later. In spite of our approximated treatments, it captures one very important, and unexplored effect: the stochastic GWB from a cosmologically motivated population of *high-eccentricity SMBH inspirals.* When dynamical friction acts on a satellite SMBH with $q \ll 1$ in a Keplerian potential and a relatively flat density profile, the orbit becomes increasingly eccentric (Antonini & Merritt, 2012). In some portions of the parameter space, the final parsec problem can be self-consistently bypassed by eccentric dynamical friction effects. Specifically, for $a_{\rm h} \ll r_{\rm in}$ and sufficiently small q, the secondary's pericenter will decrease much more rapidly than its apocenter, allowing it to bypass the final parsec problem altogether by using apocentric interactions as a sink for angular momentum at roughly fixed energy. We analyze this effect in greater detail in later sections.

Together, these two models allow us to separately explore the role of dynamical friction (FLC-model) and that of possible three-body interactions (ELC-model) on BH mergers, especially merger rates and stochastic GWB. To proceed further, it is very important to establish a proper criterion for BH mergers. Given our two limiting treatments for dynamical friction, we adopt two physically motivated, but distinct merger conditions for BH mergers. We assume that BHs merge under the following conditions:

1. When dynamical friction is not zero ($f_{\rm df} \neq 0$):
   If gravitational wave (GW) emission becomes efficient ($P_{\rm GW} > P_{\rm df}$) over multiple orbits, the binary is declared as a merged BH when the decay time due to GW emissions is shorter than the dynamical time scale $t_{\rm dyn}$.

[25] This model name, as well as the assumptions behind this model, may be overly idealized. However, our strategy here is to anchor our two models as extreme, but physically possible end limits for BH merger scenarios.

2. When dynamical friction is zero ($f_{\rm df} = 0$):
   If the decay time due to GW is shorter than the time left until the next galaxy merger and $t_{\rm dyn}$, the binary is declared as a merged BH.

3. For either $f_{\rm df} = 0$ or $f_{\rm df} \neq 0$:
   If the Schwarzschild radii of two BHs overlap, the binary immediately merges. Simply: $r < r_{\rm sch,1} + r_{\rm sch,2}$, where r is the separation of two BHs and $r_{\rm sch}$ is the BH Schwarzschild radius.

The decay time due to GW emissions is evaluated as $|a/\dot{a}_{\rm GW}|$ using Equation (5.6) in Peters (1964). The code computes, and updates at every time step, the decay time until merger. In condition 1, P represents the dimensionless dissipative power and time scale for each force, defined as $P_{\rm GW,df} = \boldsymbol{f}_{\rm GW,df} \cdot \boldsymbol{v}(E_{\rm b}/t_{\rm dyn})^{-1}$, where $E_{\rm b}$ is the orbital binding energy. In the simulations, whether BHs would merge in the FLC-model is mostly decided by condition 1, while in the ELC-model, by condition 2. Condition 3 may not even be relevant when two BHs form binaries and merge without the help of other BHs (likely in the FLC-model), but we include it to account for possible collision events in chaotic multi−BH interactions (the ELC-model).

5.3.5 Gravitational wave recoils and remnant masses

When two SMBHs merge, the remnant BH gets a kick due to anisotropic emission of gravitational waves (Bekenstein, 1973; Fitchett & Detweiler, 1984; Favata et al., 2004). Recent numerical simulations of general relativity have confirmed that the recoil velocities could be as large as galactic escape velocities depending on progenitor spins and mass ratios (Campanelli et al., 2007a,b; Lousto et al., 2010; Lousto & Zlochower, 2011). For such large kicks (up to ~ 5000 km s^{-1}), the remnant BH could escape to infinity or end up orbiting in the outskirts of the halo. If the kicks are not large enough to completely eject the remnant BH, the BH may return to the core regions after temporarily being ejected, taking part in interactions again with other BHs.

We implement the effects of the recoil kick in the simulations and take into account the mass loss to gravitational radiation for the remnant BH using the analytic formulae with the best-fit values given in Lousto et al. (2010), with random spin orientations and dimensionless spin magnitudes chosen randomly between 0 and 0.9. We provide the detailed expressions and prescriptions used in this study in Appendix A.

5.4 The equations of motion

Using a few-body code (see Ryu, Tanaka, & Perna 2016; Ryu et al. 2017a for code details), the equations of motion and mass growth for each SMBH embedded in the evolving galaxies are integrated. The motion of the BHs is determined by the following forces: (i) $\boldsymbol{a}_{\rm N} + \boldsymbol{a}_{\rm PN}$: their mutual gravitational attraction including post-Newtonian terms up to 2.5th order, (ii) $\boldsymbol{a}_{\rm df}$: dynamical friction from the surrounding medium (stars+ DM), (iii) $\boldsymbol{a}_{\rm bg}$: the gravitational pull of the background matter (stars+ DM) and (iv) $\boldsymbol{a}_{\rm mg}$: the deceleration due to BH mass increase with momentum conserved. The resulting equation of motion for the $i^{\rm th}$ BH includes

the sum of the five forces:

$$\boldsymbol{a}_i = \boldsymbol{a}_{\mathrm{N},i} + \boldsymbol{a}_{\mathrm{PN},i} + \boldsymbol{a}_{\mathrm{df},i} + \boldsymbol{a}_{\mathrm{bg},i} + \boldsymbol{a}_{\mathrm{mg},i} \tag{95}$$

Given the solutions of the equation of motion at every time step, we update the positions and velocities for each BH and the evolution of galaxy potentials. We next describe each contribution in detail.

1. *Mutual gravitational forces between BHs*
   We calculate the standard Newtonian gravitational force $\boldsymbol{a}_{\mathrm{N}}$ as well as post-Newtonian terms $\boldsymbol{a}_{\mathrm{PN}}$,

$$\begin{aligned}\boldsymbol{a}_{\mathrm{gr}} &= \boldsymbol{a}_{\mathrm{N},i} + \boldsymbol{a}_{\mathrm{PN},i} \\ &= -\sum_{j\neq i} G\ M_{\mathrm{BH},j}\ \frac{\partial\ \Phi(r_{ij})}{\partial\ r_{ij}}\ \frac{\boldsymbol{r}_i - \mathbf{r}_j}{r_{ij}} \\ &+ \boldsymbol{a}_{\mathrm{1PN},i} + \boldsymbol{a}_{\mathrm{2PN},i} + \boldsymbol{a}_{\mathrm{2.5PN},i},\end{aligned} \tag{96}$$

   where G is the gravitational constant, Φ is the pairwise gravitational potential, $\mathbf{r}_i$ is the displacement of the i^{th} BH from the center of the host galaxy, and $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$. In our numerical implementation, we adopt the Plummer softening kernel (e.g. Binney & Tremaine, 1987) with softening length equivalent to the Schwarzschild radius for a 100 $\mathrm{M}_\odot$ BH.

   We include post-Newtonian terms $\boldsymbol{a}_{\mathrm{PN}}$ up to order 2.5, which account for the loss of orbital energy and angular momentum via gravitational waves, but do not account for spin-orbit or spin-spin coupling. The full expressions for these terms can be found in, e.g., (Kupi et al., 2006).

2. *Dynamical friction from background matter*
   When an object moves through a medium, it induces an overdensity of the medium, or wake, behind it. The gravitational force due to the overdense region acts as a dissipative drag on the object's motion. In this study, we consider dynamical friction due to both DM and stars.

   For the DM contribution, we adopt the standard Chandrasekhar formula Binney & Tremaine 1987,

$$\boldsymbol{a}_{\mathrm{df},i} = -4\boldsymbol{\pi}\ \ln\Lambda\ f(X_i)\ \frac{G^2 M_{\mathrm{BH},i}}{v_i^3}\ \rho(r_i)\ \boldsymbol{v}_i, \tag{97}$$

   with

$$f(X_i) \equiv \mathrm{erf}(X_i) - \frac{2}{\sqrt{\boldsymbol{\pi}}}\ X_i\ \exp\left(-X_i^2\right), \tag{98}$$

   where $X_i \equiv v_i/(\sqrt{2}\sigma_v)$. We use the circular velocity, defined as $\sqrt{G[M_{\mathrm{en}}(r \leq r_i) + M_{\mathrm{BH}}(r \leq r_i)]/r_i}$ for σ_v. We do not include the contribution of stars bound to infalling BHs in estimating dynamical friction. Again, $M_{\mathrm{en}}(r \leq r_i)$ is the enclosed mass (DM+stars) and $M_{\mathrm{BH}}(r \leq r_i)$ is the total mass of BHs (including the

i-th BH itself) inside $r = r_i$. The expression for the enclosed mass of stars is given in Stone & Ostriker (2015). We use $\ln \Lambda = 5$ (Spinnato et al., 2003; Merritt, 2006) and we take the sum of local densities of stars and DM for ρ, namely, $\rho = \rho_\star + \rho_{\rm DM}$, at the location of the i-th BH.

3. *Gravitational force of the background matter*
The background stars and DM exert an additional gravitational force on the BHs. Because we assume a spherically symmetric density profile, this force points toward the centre of the potential. It can be expressed as

$$\boldsymbol{a}_{{\rm bg},i} = -\frac{G\ M_{{\rm en},i}(r \leq r_i)}{r_i^3}\mathbf{r}_i, \tag{99}$$

where $\mathbf{r}_i$ is a vector pointing from the centre of the galaxy to the i-th BH.

4. *Deceleration due to mass growth*
We take into account the decrease in velocity due to mass growth (see §5.3.2). Assuming BHs grow in mass in a spherically symmetric fashion, the $i-$th BH decelerates through conservation of linear momentum,

$$\boldsymbol{a}_{{\rm mg},i} = -\frac{M'_{\rm BH,i} - M_{\rm BH,i}}{M'_{\rm BH,i}\Delta t}\boldsymbol{v}_i\,, \tag{100}$$

where M'_i is the increased mass estimated using Equation 92, and Δt is the time step.

In summary, our simulations display several noticeable features: (i) we follow the merger history of SMBH host galaxies as extracted from cosmological N-body simulations for $0 < z < 4$, across a wide range of merger mass ratios, $10^{-4} < q_\star < 1$; (ii) we take into account the evolution of the galactic potential (star+DM) in both physical size and depth as a result of both galactic mass growth and core scouring from SMBH binaries; (iii) we explore two different models, the empty and full loss cones. These two extreme assumptions plausibly bracket the true evolution of binary BHs close to their merger. In addition, they allow us to clearly isolate the importance of multi-body BH interactions between BHs at coalescence. We show this by estimating the BH merger rates and the GWB independently for the two models.

5.5 Results

In this section we present the binary lifetimes of merged BHs and their merger rates for the two models (FLC and ELC). Additionally, given the merger rates, we infer the characteristic GW amplitude $h_{\rm c}$. Given the eccentricities found in our simulations, we show how $h_{\rm c}$ for eccentric orbits deviates from that for circular orbits.

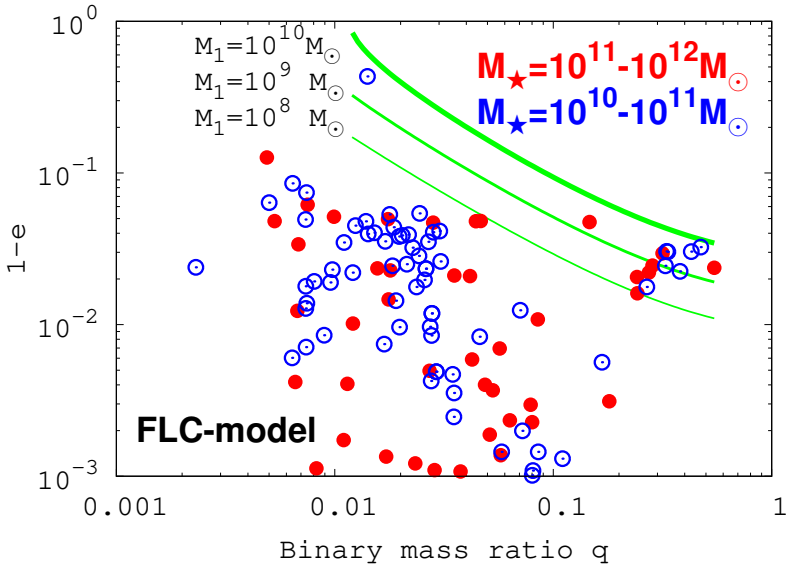

Figure 25: The distribution of the eccentricities of merged binaries in the FLC-model as a function of the binary mass ratio q when energy loss by GWs becomes dominant. We use different colors to distinguish between merged binaries in the galaxies of $M_\star = 10^{11} - 10^{12}$ $M_\odot$ (red) and in $M_\star = 10^{10} - 10^{11}$ $M_\odot$ (blue). The eccentricities are quite high ($e > 0.9$). Green curves show analytic limits for the validity of our approach (Equation 101), which estimates dynamical friction in the absence of stellar scattering. High-q mergers above these curves are not treated self-consistently by our FLC-model, but the majority of (low-q) mergers, which lie below these curves, are.

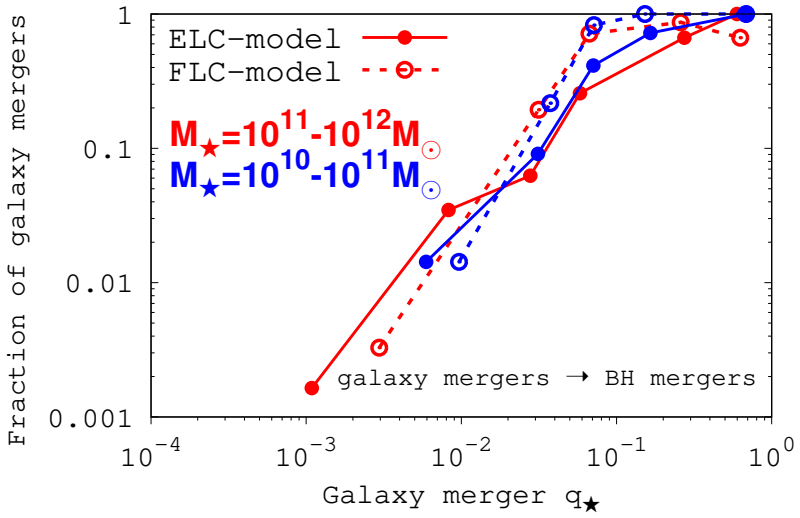

Figure 26: The fraction of galaxy mergers, as a function of their mass ratio $q_\star$, for which the central BHs have merged over our entire merger trees. For example, the number fraction of 1 (0) means that the galaxy merger always (never) leads to coalescence of the central BHs.

Table 12: Overview of BH coalescence events for the FLC- and ELC- model for the host galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ and $M_\star = 10^{11} - 10^{12}$ $M_\odot$. From top to bottom: The mass of sampled galaxies, the total number of BH coalescences, the average mass of merged binary BHs in unit of $M_\odot$ and $M_\star$, the number fraction of BH coalescences in host galaxies with one significant merger ($N_\star = 1$) with $q_\star \geq 0.01$ and that in host galaxies with multiple significant mergers ($N_\star \geq 2$) with $q_\star \geq 0.01$. Notice that only 1% of BH coalescences occur in host galaxies with $M_\star = 10^{10} - 10^{11}$ $M_\odot$ experiencing only one significant merger ($N_\star = 1, q_\star \geq 0.01$). This is because the number of such galaxies is small (See Figure 22 and Table 11) and binary formation and BH mergers are less likely to happen.

Galaxy mass $M_\star$	$10^{10} - 10^{11}$ $M_\odot$		$10^{11} - 10^{12}$ $M_\odot$	
loss cone models	FLC-model	ELC-model	FLC-model	ELC-model
Total number of BH coalescences	140	77	76	35
Average total mass of merged BHs [10^8 $M_\odot$]	2.2	3.2	15	17
Average total mass of merged BHs [in unit of $10^{-3} M_\star$]	4.4	5.0	4.5	5.1
Number fraction of BH coalescences in host galaxies ($N_\star = 1$, $q_\star \geq 0.01$)	1%	1%	17%	8%
Number fraction of BH coalescences in host galaxies ($N_\star \geq 2$, $q_\star \geq 0.01$)	99%	99%	83%	92%

5.5.1 Dynamical features

1. *FLC-model*

   In the FLC-model, the birth eccentricities of the binaries are moderate ($e \gtrsim 0.4$). These are lower than the eccentricities assigned to BHs as initial conditions (see Equation 93). This is because inspiralling BHs experience the strongest dynamical friction forces (prior to binary formation) at pericenter (near denser core region), leading to orbital circularization. Given that the density profile adopted in this study approximately follows $\rho \sim r^{-2}$ at $r_{\rm c} < r < r_{\rm h}$, this is consistent with the eccentricity evolution of BHs in an isothermal density profile decaying towards the core shown in Ryu, Tanaka, Perna, & Haiman (2016) (see their Figure 7). Once a binary forms, however, orbital eccentricities increase rapidly due to dynamical friction. At the point when GW emission becomes the dominant driver of orbital decay, eccentricities can reach up to $e > 0.99$ and semimajor axes down to $a \sim 0.01 - 1$ pc.

   We emphasize here that the evolution of the eccentricities in the FLC model likely represents the most extreme scenario of eccentricity evolution. Accounting for stellar 3-body scatterings would likely moderate the increase in eccentricity we observe. Hence, the eccentricity at which GW emission takes over may not be as high as that found in this study. Indeed, the eccentricities of compact binaries in our simulations tend to be higher than those found in some previous numerical works with large N-body simulations (e.g. Berentzen et al., 2009; Khan et al., 2011; Preto et al., 2011), even though a qualitatively similar increase in eccentricity has been seen in those studies. Since a long-term "full" loss cone in large N-body simulations cannot be easily achieved, the binary evolutions found in their studies may correspond to intermediate regimes bracketed by our two models. For example, Berentzen et al. (2009) studied the evolution of SMBH binaries, focusing on the interactions with surrounding stars. In the eccentricity evolutions shown in their examples, we can see a rapid increase right after binary formation, followed by a relatively gradual rise. This may be due to quick depletion of the initially full loss cone reported in their paper, as noted above, possibly corresponding to a regime in between our two models.

   We show in Figure 25 the distribution of the eccentricities of binaries in the FLC-model which will eventually merge, as a function of the mass ratio q. The eccentricities are evaluated at the time when GWs become more efficient. For such eccentric binaries, the decay time (Peters, 1964) is short (typically, $t_{\rm decay} < 10^8$ yr). Considering the galaxy merger time scale of ~ 1 Gyr and the long infall times for BHs to reach the core, this means coalescences of BHs may occur even before a 3$^{\rm rd}$ BH can arrive. Indeed, in almost all of our FLC-model simulations, incoming BHs which can reach the core form binaries with the central BH, and subsequently merge on a short time scale. Of course, our FLC-model orbital evolution is quite approximate in that it neglects hardening via three-body interactions with surrounding stars. This approximation is only justified in the subset of parameter space where a radializing binary orbit (inside the primary influence radius) can keep its apocenter outside the hard radius a_h. In other words, the final parsec problem can only be bypassed when $r_{\rm p,GW} > a(1-e)$ and, simultaneously, $a > a_{\rm h}$. Here $r_{\rm p,GW}$ is the maximum pericenter for which a SMBH binary will merge

in a Hubble time $t_{\rm H}$. Combining these two inequalities gives a necessary condition for this bypass to occur, which is

$$\frac{q^{3/4}}{(1+q)^{5/4}} < \frac{4\sigma_\star^2}{c}\left(\frac{85 t_{\rm H}}{3GM_{\rm BH,1}c}\right)^{1/4}(1-e^2)^{-7/8}. \quad (101)$$

Green curves illustrating this inequality are shown for different primary SMBH masses in Fig. 25. Most of the mergers we simulate are at sufficiently low mass ratio that our simulations of high-eccentricity coalescence are self-consistent. However, we caution that Equation 101 is a necessary, not a sufficient, criterion for an eccentric bypass of the final parsec problem (see also the discussion of Antonini & Merritt 2012). Whether or not an individual secondary BH can make use of this route to coalescence depends on its initial eccentricity and on the role of three-body scatterings with stars. In addition, the degree of nuclear rotation can affect whether or not they circularize or radialize (e.g Rasskazov & Merritt, 2017; Mirza et al., 2017).

If binary lifetimes are sufficiently short that BHs coalesce before another BH makes it to the core, then BH merger rates and infall timescales of incoming BHs should have an inverse correlation. Given the shorter infall times of the more massive BHs, BH merger rates should hence increase as $q_\star$ increases. We confirm this relation in Figure 26. The plot shows the fraction of galaxy mergers of mass ratio $q_\star$, for which the central BHs are able to coalesce up to $z = 0$ in our simulations. A number fraction of 1 means two BHs introduced by a galaxy merger always successfully merge whereas a fraction of 0 means they fail to merge. In the FLC-model, as $q_\star$ increases, it is more likely that BH mergers take place, and the BH merger rates can be directly related to the frequency of major galaxy mergers.

2. *ELC model*

On the other hand, in the ELC-model, the central binaries typically stall at $r \sim$ a few 10 pc at $z = 0$. This separation may be somewhat larger than generally assumed. In our simulations, due to dynamical friction, the BH binary orbits efficiently decay to near the hardening radii, which are at least on the order of a few tens of pc at low z for the very high mass BHs we consider. Under these conditions, unless a 3$^{\rm rd}$ BH approaches sufficiently close to the core, the central binaries do not merge. This means that in order for the central binaries to further decay and finally merge, multiple (at least $N \geq 2$) major mergers are necessary, so that new BHs can make it to the core rapidly and effectively interact with the central binaries. Therefore, there is a longer delay in time from binary formation to BH merger. This is clearly different from the FLC-model. As a result, coalescences of BHs preferentially take place in the host galaxies experiencing more than one major mergers. We find in our simulations that 99% of BH mergers in the ELC-model occur in such galaxies (experiencing multiple major mergers) in both mass bins of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ and $M_\star = 10^{11} - 10^{12}$ M$_\odot$. Furthermore, we see similar correlations between the galaxy merger mass ratios and the likelihood of BH mergers within the FLC-model, as shown in Figure 26. However, we note that the fraction is slightly lower for high $q_\star$ than in the FLC-model. In

the FLC-model, major galaxy mergers favorably lead to BH mergers, but because of ejections ($\sim 1-5\%$ of BHs found at $r > r_{\rm h}$ at $z = 0$) via multiple BH interactions, this is not always the case in the ELC-model.

The general picture of multi-body BH interactions in our simulations is as follows. When a third BH is orbiting far from the core region, its orbit is governed by dynamical friction and the galactic potential. Every time the intruder BH gets sufficiently close to the central binary at the pericenter of its orbit around the galactic potential, it goes through multiple gravitational slingshots with the central binary (typically, its apocenter remains outside the core). The intruder BH gains energy at the pericenter via the slingshot mechanism, and loses energy outside the core region via dynamical friction. In this case, the background potential when the ejected BH is outside the core can additionally provide more chances to return back for another slingshot (Ryu et al., 2017a). This appears to make the intruder BH linger a little bit longer before its apocenter completely falls into the core. These ejections confined in $r < r_{\rm h}$ are not always efficient at decaying the central binary orbit [26], but initially wide binaries can benefit from these slingshots, becoming hardened to some extent.

Finally, when the three BHs become bound, they either go though chaotic interactions followed by ejections, or form a hierarchical triple. Due to the gravitational pull from the third BH, the central binaries are usually located off-center when the triples form. The central binaries go through this course of interaction, similarly described by Hoffman & Loeb (2007), one or even more times before they finally merge. We find that it is less likely for ejected BHs to return and manage to merge. Typically ejected BHs are the less massive ones, hence they tend to be easily ejected again even though they can make it to the core. Additionally, we find that escapes of all three BHs are rare (also similarly to Hoffman & Loeb 2007). Even for that case, cores empty of BHs are transient, and are rapidly re-filled with other BHs from minor mergers or the ejected BHs themselves when they return. In our simulations, BH binaries merge in hierarchical triples and due to strong binary-single BH interactions (see also Iwasawa et al. 2006). However, the majority of BH mergers occur when they are in hierarchical triples.

5.5.2 Merger efficiencies and binary lifetimes

In order to highlight the differences between the two models, we provide the average lifetimes of coalescing binaries in Figure 27 as a function of z (*upper* panel) and the binary mass ratio q (*lower* panel). We define the lifetime of a binary as the time from binary formation to coalescence. In the ELC−model, three-body interactions can cause the ionization of existing binaries. In this case we estimate the lifetime as the time between when a binary forms and when it merges, for the subset of binaries that avoid ionization. In both panels, as expected, the lifetimes of the merged binaries in the ELC-model (≥ 1 Gyr) are longer than those in the FLC-model (≤ 1 Gyr).

[26] For a hard binary (primary mass of $m_1 = 10^8$ M$_\odot$, $q = 0.1$ and $a_{\rm h} \simeq 1$ pc) with orbital energy $E_{\rm b,hard}$, the energy taken from the binary by a light BH of mass m_3 approaching with velocity $v = \sigma_\star$ and subsequently ejected at $v < v_{\rm esc}(r = r_{\rm h})$ (the escape velocity at $r = r_{\rm h}$) is $|\Delta E_{\rm b,hard}/E_{\rm b,hard}| \simeq 0.003 - 0.3$ for $m_3/m_1 = 0.001 - 0.1$.

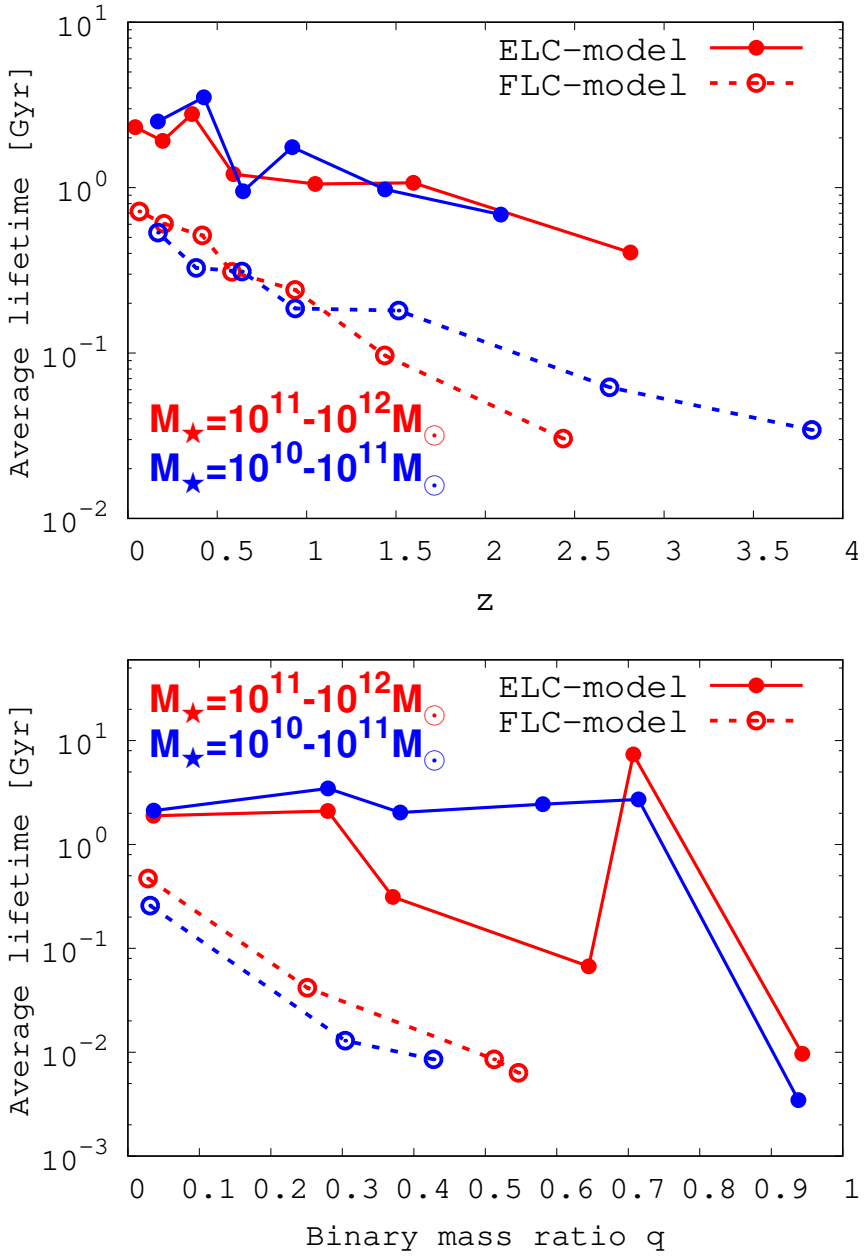

Figure 27: The average lifetimes of merged binaries as a function of z (*upper* panel) and the binary mass ratio q (*bottom* panel) for the galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ (blue lines) and $M_\star = 10^{11} - 10^{12}$ M$_\odot$ (red lines). We use solid (dotted) lines to represent the ELC- (FLC-) model. We define the lifetimes of binaries as the time from binary formation to coalescence.

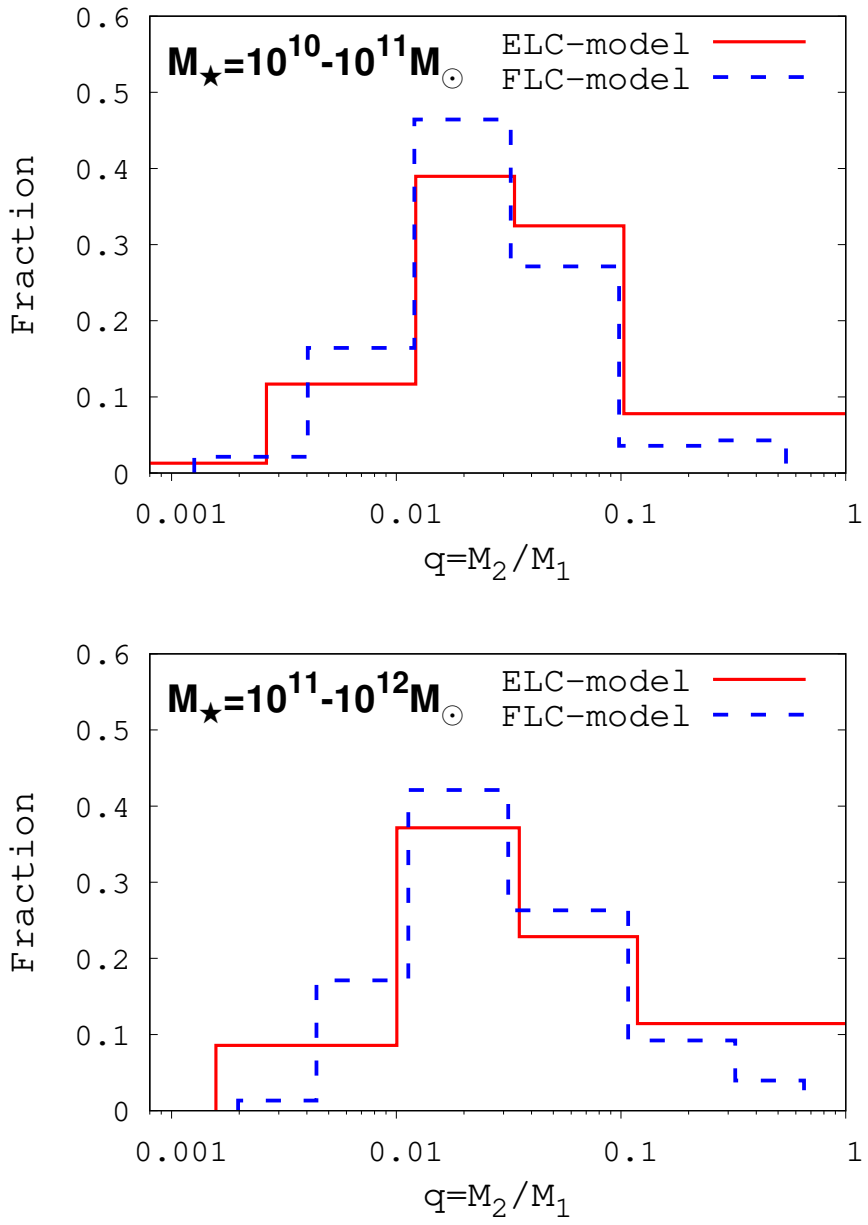

Figure 28: The relative fraction of merged central BH binaries as a function q (in logarithmic intervals) in host galaxies of masses $M_\star = 10^{10} - 10^{11}$ $M_\odot$ (*left* panel) and $M_\star = 10^{11} - 10^{12}$ $M_\odot$ (*right* panel). Red (blue) solid lines refer to the ELC- (FLC-) model. It is normalized such that the sum of the fractions is unity.

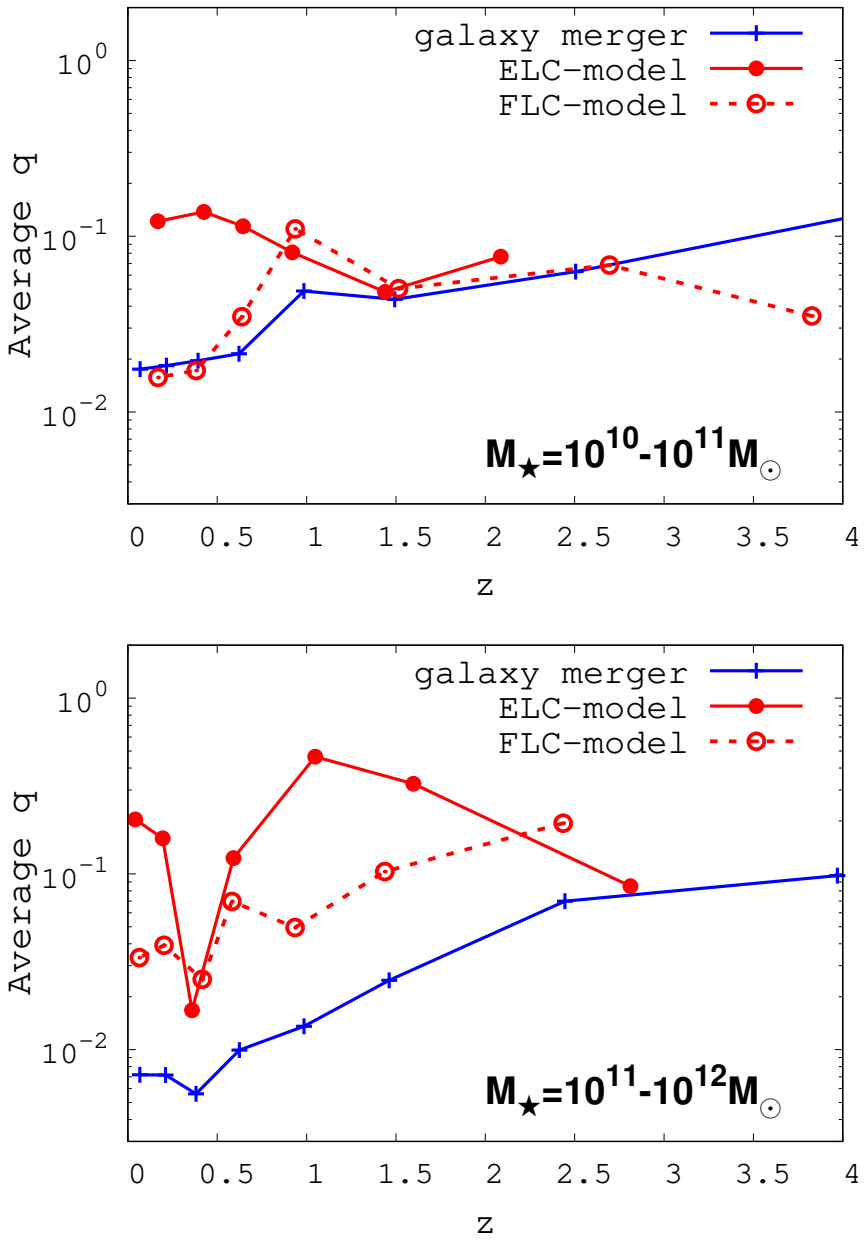

Figure 29: The average q in each Gyr from $z = 3.5$ to $z = 0$, along with the average galaxy merger ratio $q_\star$ (same lines as in the *middle* panels in Figure 22, but each line separately drawn in each panel).

The lifetimes for the FLC-model that we find are relatively short compared to the ones reported by Kelley et al. (2017). Besides different models and different prescriptions for binary decay mechanisms, this difference may be primarily due to highly eccentric binary orbits and the assumption of efficient decay due to dynamical friction at all times. In particular, in the *upper* panel, as z decreases, the lifetimes become longer for both models. However, such longer times may be due to different reasons in each model. In the FLC-model, we can understand this as a result of galaxy mergers of smaller q at lower z (see the *middle* panel in Figure 22), hence binaries with lower $q_\star$. Remember that the dynamical friction timescale for a tightly bound binary is roughly estimated as $t_{\rm df} = E_{\rm b}/(f_{\rm df} \cdot v) \propto M_{\rm BH,1}^{1/2} q^{-1}$. This can be also found in the *bottom* panel, which shows that the lifetimes rise as q declines. In the ELC-model, on the other hand, the longer lifetimes may be attributed to mainly two reasons: 1) as galaxy mergers occur with smaller $q_\star$ the central binaries have to wait for a longer time until new BHs fall into the core (or longer infall times of less massive BHs); and 2) it is harder for the central binaries to be ionized or to get hardened via three-body interactions. Interestingly, differently than in the FLC-model, the dependence on the mass ratio q is weakened (even flat for $q < 0.3$) as the central binaries go through chaotic interactions with other BHs, followed by ionization and exchange in binary members.

Because of such differences between the two models, we find different statistical properties of the merged BH binaries including their merger rates and mass ratios. This is the subject of the next section.

5.6 Coalescence of BHs - BH merger rate and mass ratio

In this section we focus on a detailed analysis of the statistical distributions of BH mergers, such as merger rates, mass ratios and their evolution as a function of z. We provide an overview of BH coalescence events for the FLC- and ELC- models in Table 12.

5.6.1 Mass ratios and chirp mass of coalescing BH binaries

In Figure 28 we present the number fraction of merged central BH binaries as a function of q (in logarithmic intervals) in host galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ (*left* panel) and $M_\star = 10^{11} - 10^{12}$ M$_\odot$ (*right* panel). A noticeable difference between the ELC and the FLC-model is that BH mergers with larger mass ratios are more common in the ELC-model (see longer high-q tails for the ELC-model in both galaxies). The reason for this is likely the nature of three-body interactions, i.e., less massive objects being easily ejected, leaving behind more massive binaries (Valtonen & Karttunen, 2006). This trend is more pronounced in the host galaxies of $M_\star = 10^{11} - 10^{12}$ M$_\odot$ (*right* panel). Considering more frequent major mergers (see Table 11) as well as higher $q_\star$ (see Figures 22), the BH merger ratios in such galaxies for the FLC-model and ELC-model are generally high. However, for the galaxies of $M_\star = 10^{11} - 10^{12}$ M$_\odot$, the number of host galaxies going through a single major merger and multiple major mergers are comparable (ratio of $\sim 3:4$ in Table 11). This means that BH mergers in the ELC-model more "selectively" occur in the galaxies experiencing multiple major mergers. Even though the merger rates are low (see Figure 32), this can possibly lead to a shift to higher q.

Such enhancement of higher q (or "selective mergers" in more massive galaxies) for the ELC-model can also be found in Figure 29. In this figure we show the average q for every Gyr from $z = 4$ to $z = 0$ along with the average galaxy merger ratio $q_\star$. As explained above, typically the mass ratios for the ELC-model are higher than for the FLC-model. However, comparing with the galaxy merger ratios, the difference becomes noticeable. For galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ (*left* panel), the BH merger mass ratios q are quite moderately following the line for the galaxy merger mass ratio $q_\star$. For those of $M_\star = 10^{11} - 10^{12}$ M$_\odot$ (*right* panel), however, the lines for q are always positioned above that for $q_\star$, and q for the ELC-model is generally higher than that for the FLC-model.

As a consequence of three-body interactions, the chirp mass is higher for BH mergers in the ELC-model. We present these effects in Figures 30 and 31. Figure 30 shows the fraction of ejected BHs as a function of mass ratio q in the ELC-model. Here, q labeled "ejected BHs" refers to the mass ratio of ejected BHs to the central BHs during 3-body interactions. As a comparison, we also depict the lines corresponding to the merged binaries shown in Figure 28. We can see a higher fraction of ejected BHs with smaller q for galaxies in both mass bins. This implies that less massive BHs are more likely to be ejected, resulting in more massive binaries retained in the core regions. Additionally, a comparison between the two panels shows that the mass ratios of ejected BHs to the central BHs in larger galaxies (*left* panel) are lower than those in smaller galaxies (*right* panel). Therefore, given the central binary masses required by the $M - \sigma$ relation (i.e., the average mass of merged binaries $\sim 4.5 \times 10^{-3} M_\star$ in Table 12) and the larger mass ratios, the chirp mass for the ELC-model also becomes higher for galaxies in both mass bins as found in Figure 31. The shaded regions indicate 68% of BH mergers at a given redshift. The lines for the average values and those demarcating the shaded regions share the same line types, but slightly thinner.

5.6.2 BH merger rate

We present in Figure 32 and Figure 33 two different realizations of the BH merger rates as a function of z. Figure 32 shows the merger counts per central BH/galaxy averaged over every Gyr, or $\Delta N/\Delta t$ for the host galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ (*left* panel) and $M_\star = 10^{11} - 10^{11}$ M$_\odot$ (*right* panel), with a reference line corresponding to $\Delta N/\Delta t = 0.1$. There are a few noticeable features seen in both panels as follows: (i) the BH coalescence rates for the FLC-model are higher for galaxies in both mass bins than those for the ELC-model. This is also seen in Figure 33. This is expected given the longer lifetimes of BH binaries in the ELC-model, possibly leading to ionizations of binaries as well as ejections of BHs; (ii) the merger rates are higher for BHs in less massive host galaxies (*left* panel). Notice that the BH merger rates for more massive galaxies are always below the reference line; but the differences in the BH merger rates between the galaxies get smaller as z decreases. Finally, the rates tend to converge to 10^{-2} Gyr$^{-1} < \Delta N/\Delta t < 10^{-1}$ Gyr^{-1} at $z \simeq 0$. The rate at $z = 0$ is consistent with what has been assumed as a present-day merger rate for a single object in Jaffe & Backer (2003). (iii) Comparing the BH merger rates with the galaxy merger rates, the BH coalescence rates are smaller than the galaxy merger rates by a factor of $3 - 20$ depending on the model and redshift. As shown in Figure 26, every galaxy merger with a small mass ratio does not always lead to a BH merger. BHs, which either never fall into the core or are ejected, are left orbiting outside the core regions. If one only considers

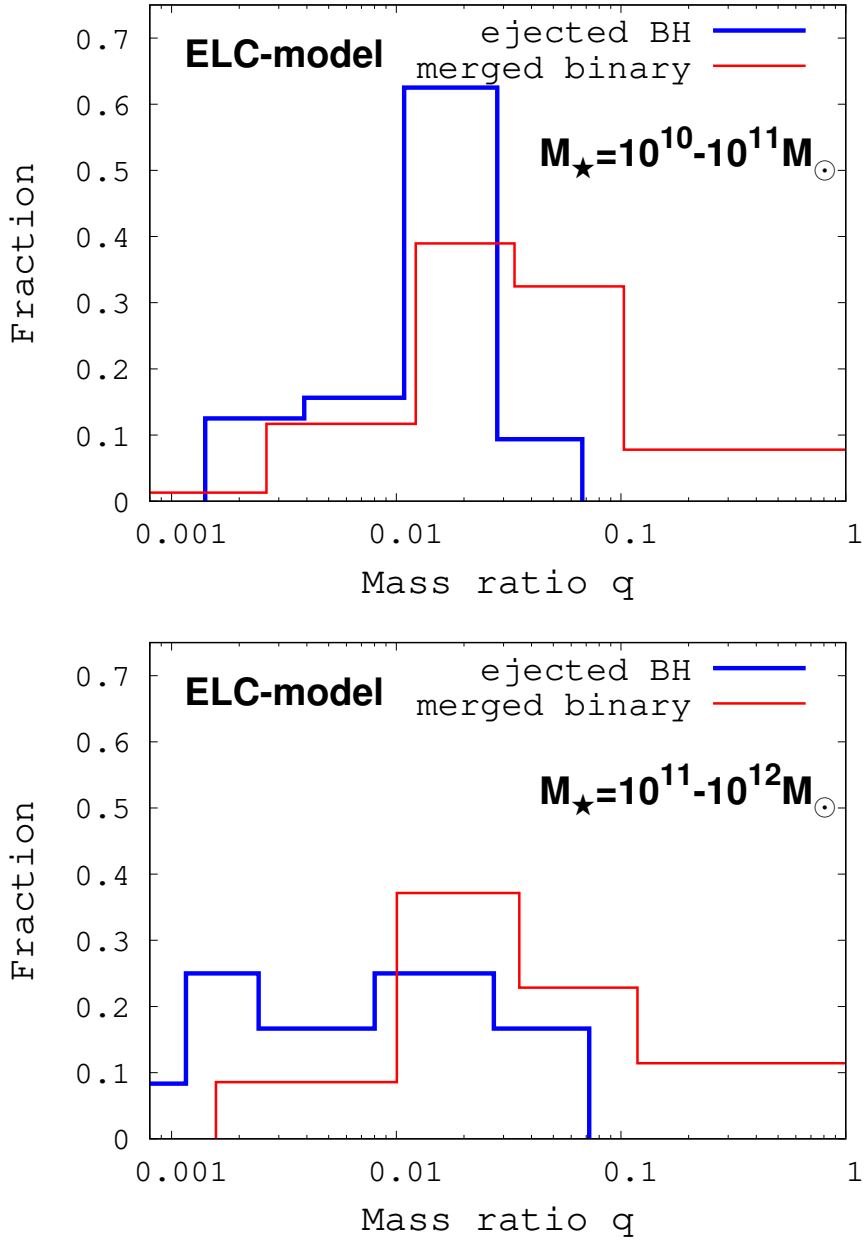

Figure 30: The relative fraction of ejected BHs (thick blue solid line) as a function of mass ratio q in the ELC-model. This is normalized so that the sum of the fractions is unity. Here, q of the "ejected BHs" is defined as the mass ratio between ejected BHs and central BHs during 3-body interactions. As a comparison, we also depict the lines (thin red solid) corresponding to the merged binaries shown in Figure 28.

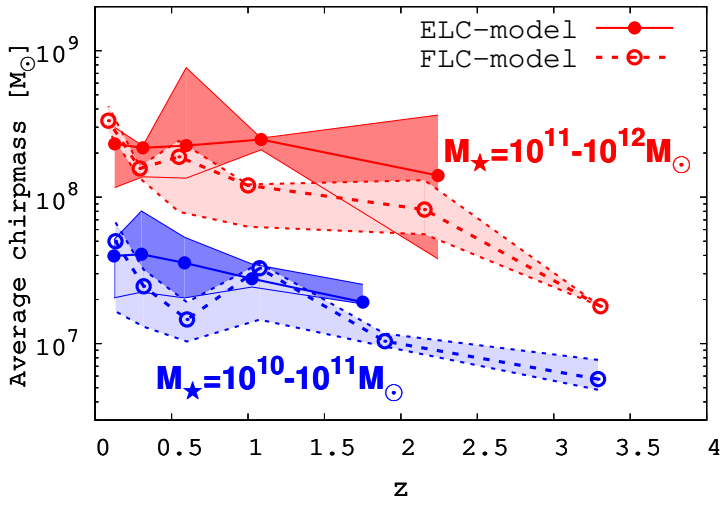

Figure 31: The redshift evolution of the average chirp mass for merged BHs in galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ (blue lines) and $M_\star = 10^{11} - 10^{12}$ M$_\odot$ (red lines). The dotted lines represent the FLC-model and the solid lines the ELC-model. The shaded regions indicate 68% of BH mergers at a given redshift. We use same line types for the average values (lines with circles) and the (slightly thinner) lines running along with the boundaries of the shaded regions.

major mergers ($q_\star > 0.1$), then as indicated in Figure 26, the differences should be smaller. However, such differences should be considered for studies including both minor and major mergers.

In Figure 33, we also show the merger rates of BHs and galaxies in two different units. In the *left* panel, we show the number of BH/galaxy mergers per unit redshift per comoving volume V_c, or $d^2N/dzdV_c$. For this, we take for V_c the size of the computation box in the Milli-Millennium simulation ($V_c \simeq 6.28 \times 10^5$ Mpc^3). It is clear that the merger rates are rising towards lower $z < 1.5$ (as those for galaxies) except for the rate of the galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ for the FLC-model, which remain roughly flat. The counts for all models reach up to $d^2N/dzdV_c \sim 2\times10^{-4}$ Mpc^{-3} for more massive galaxies and $d^2N/dzdV_c \sim 3 \times 10^{-3}$ Mpc^{-3} for less massive galaxies at $z \simeq 0$. This is attributed to the tendency for a larger number of BHs to accumulate in the core region at $z \sim 0$. Even smaller BHs (with longer decay times) can have enough time to decay to the core regions, increasing the chances of BH mergers in both models. Additionally, given the high merger rates for lower-mass galaxies, and especially the higher mass ratios in the ELC-model, we can expect that the contribution of BH mergers in lower-mass galaxies to the GWB is not negligible (see Figure 35).

In the *right* panel, the number of BHs/galaxy mergers per unit time per unit redshift, or $d^2N/dzdt$, is presented. This represents the detectable merger rate that originates from a comoving shell in redshift (corresponding to the comoving volume in the *left* panel). For the conversion between the merger rates in the *left* in the *right* panels, we use equation (4) in Menou et al. (2001). The same line colors and line types are used as in the *left* panel. Also note that, for a clearer view, we further draw on a logarithmic scale the lines for the BH

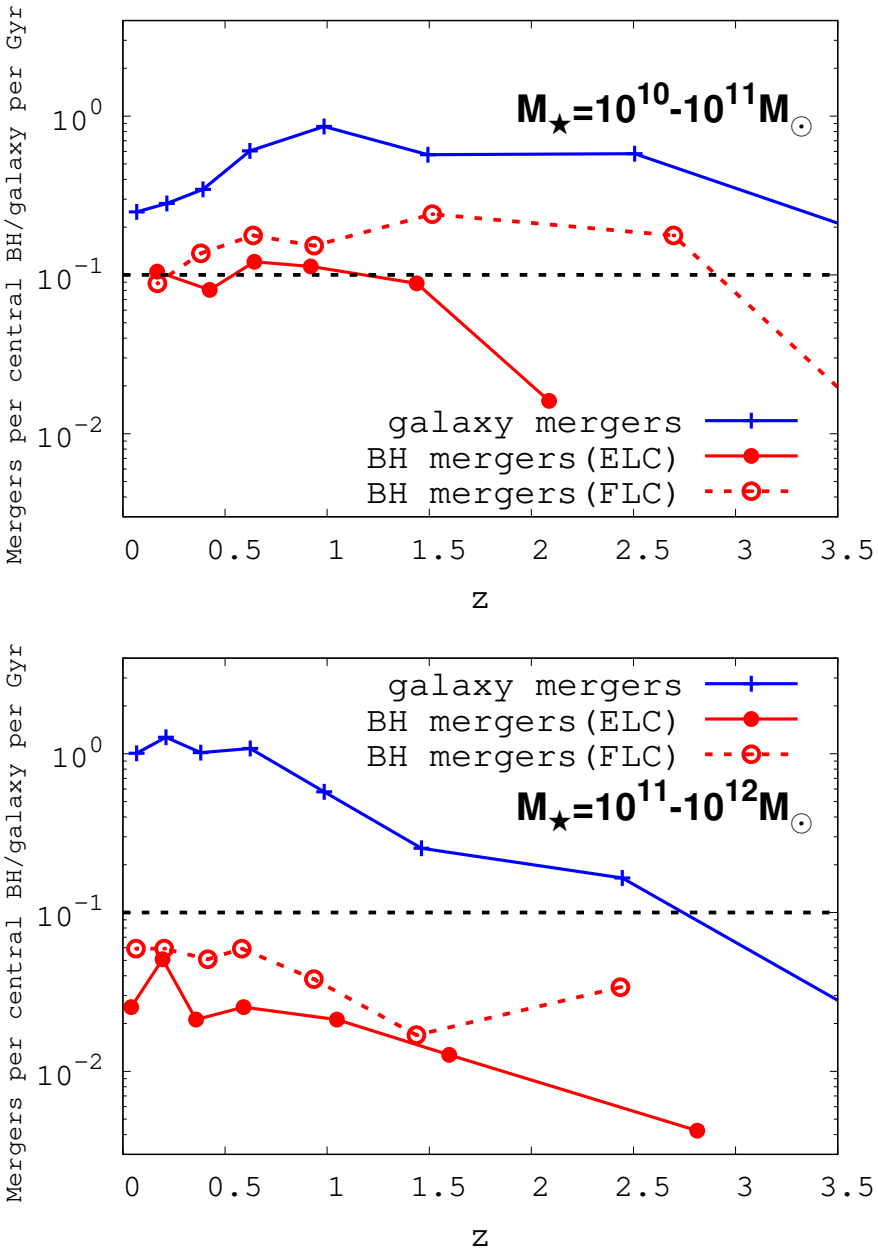

Figure 32: The merger counts per central BH (red lines)/galaxy (blue line) averaged per Gyr, or $\Delta N/\Delta t$ for the host galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ (*left* panel) and $M_\star = 10^{11} - 10^{12}$ $M_\odot$ (*right* panel). The blue solid line with crosses indicates the merger counts for the host galaxies (same lines with the thickest lines as in the *bottom* panel of Figure 22). We adopt the solid line with solid squares for the ELC-model and the dotted line with hollow squares for the FLC-model. For an easy comparison, we additionally depict a reference line (black dotted line) corresponding to $\Delta N/\Delta t = 0.1$.

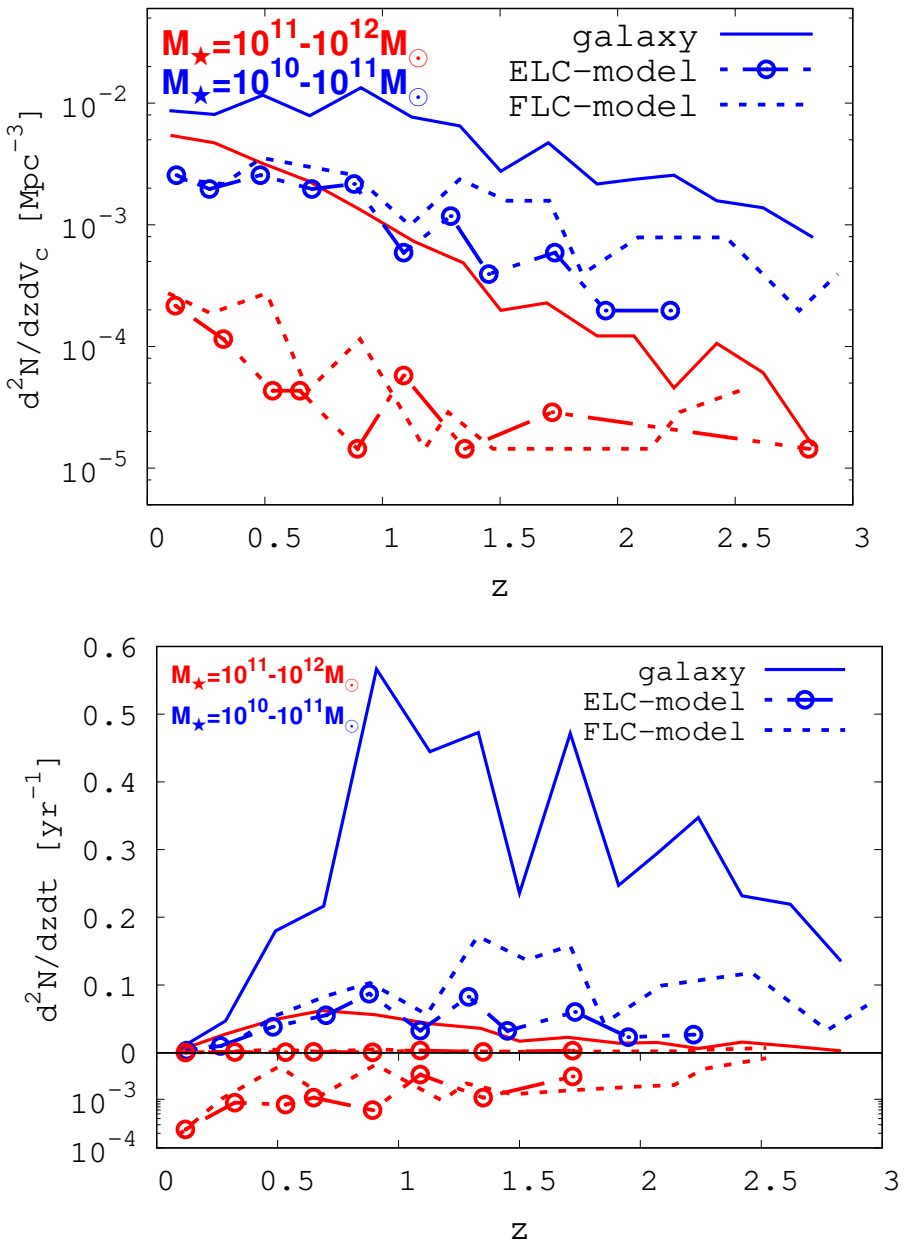

Figure 33: The merger rates of BHs and galaxies in two different units. *Left* panel: the number of BH mergers per unit redshift per comoving volume $V_{\rm c}$, or $d^2N/dzdV_{\rm c}$, for the FLC- (dotted lines) and the ELC- models (dot-dashed lines with circle). Here, we take for $V_{\rm c}$ the size of the computation box in the Milli-Millennium simulation. *Right* panel: the number of BH/galaxy mergers per unit time per unit redshift, or $d^2N/dzdt$. We use equation (4) in Menou et al. (2001) for the unit conversion between the merger rates in the two panels. The same line colors and types are adopted as in the *left* panel. In the bottom panel, for clarity, we further draw on a logarithmic scale the lines for the BH merger rates in the galaxies of $M_\star = 10^{11} - 10^{12}$ M$_\odot$.

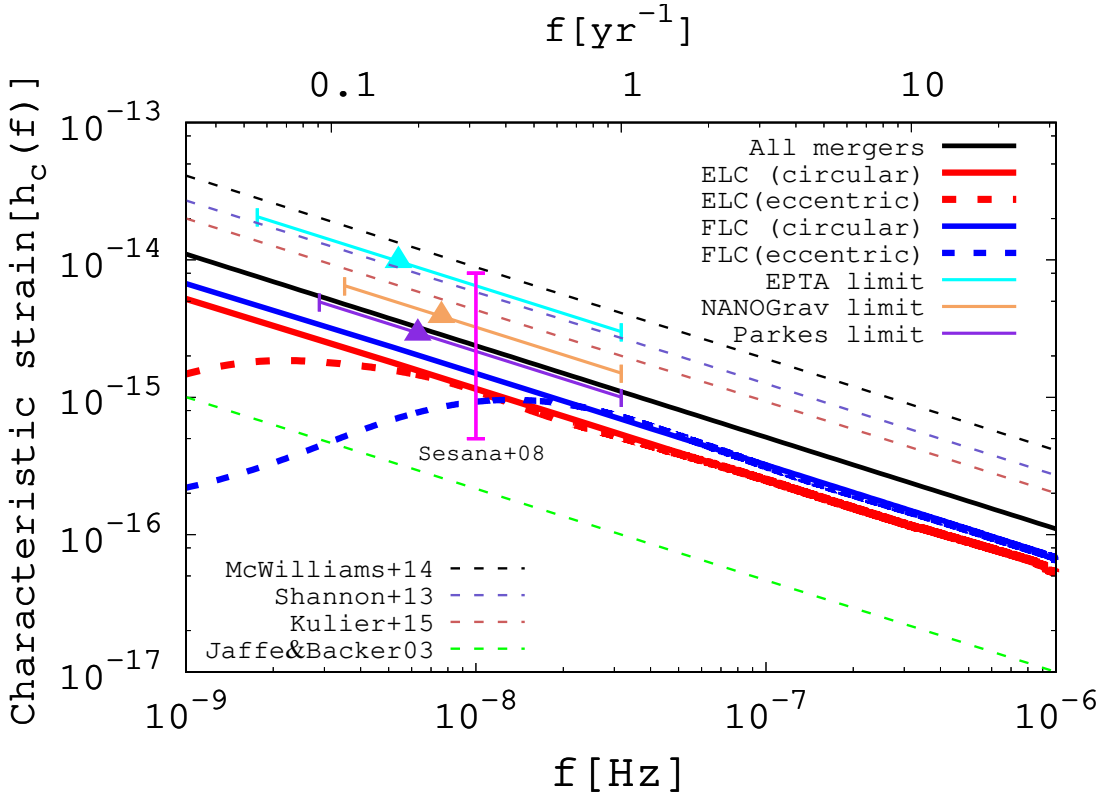

Figure 34: The characteristic strain h_c for the FLC-model (thick blue solid line) and the ELC-model (thick red solid line). The black solid line (labeled "All mergers") above both models indicates the strain assuming all galaxy mergers lead to BH mergers given the sampled galaxy merger trees (see more details in §5.7.3). We additionally indicate the upper limit in each experiment at its peak sensitivity (triangles), and we extrapolate this limit to other frequencies assuming a power law of $f^{-2/3}$ within the frequency range of $1/T < f < 1$ yr^{-1}, where T is the total observing time. The dotted lines refer to the PTA estimates from other studies. We estimate $A = 0.70 \times 10^{-15}$ for the FLC-model and $A = 0.53 \times 10^{-15}$ for the ELC-model. The curved dotted lines indicate the deviation due to eccentric orbits. The line colors are shared with those for the circular orbit case (thick blue/red solid lines).

merger rates in the galaxies of $M_\star = 10^{11} - 10^{12}$ $M_\odot$ (bottom box). The BH merger event rate is $0.1 - 0.2$ per yr at $z \simeq 1 - 2$ and $10^{-4} - 10^{-2}$ per yr for the galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ and $M_\star = 10^{11} - 10^{12}$ $M_\odot$, respectively.

In the next section, we use the BH merger rates in our models to estimate the amplitude and spectrum of the stochastic GWB.

5.7 Discussion

5.7.1 Stochastic GW background - GW from circular orbits

Over observing times of a few years to a few months, binary supermassive BHs are one of the most promising astrophysical sources of GWs in the nHz frequency band accessible to pulsar timing arrays (PTAs). The characteristic strain $h_c(f)$ of the GW signal from a circular binary can be calculated as follows (Phinney, 2001; Sesana et al., 2008),

$$h_c^2(f) = \frac{4G}{\pi f^2 c^2} \int_0^\infty dz \int_0^\infty d\mathcal{M} \frac{d^2n}{dz d\mathcal{M}} \frac{1}{1+z} \frac{dE_{\rm gw}(\mathcal{M})}{d\ln f_{\rm r}}\,, \tag{102}$$

where f is the observed frequency and $\mathcal{M}$ is the chirp mass, defined as $\mathcal{M} = (M_{\rm BH,1} M_{\rm BH,2})^{3/5}/(M_{\rm BH,1} + M_{\rm BH,2})^{1/5}$. Here f is related to the rest-frame frequency $f_{\rm r}$ and the Keplerian orbital frequency $f_{\rm orb}$ such that $f(1+z) = f_{\rm r} = 2 f_{\rm orb}$. $E_{\rm gw}$ is the energy emitted in GWs. $d^2n/dzd\mathcal{M}$ represents the differential merger rate density (i.e., the number of BH mergers per comoving volume) of SMBH binaries per unit redshift per unit chirp mass. It is easily shown that the strain scales as $f^{-2/3}$ (Phinney, 2001) and is usually described in terms of A (Jenet et al., 2006),

$$h_c(f) = A \left(\frac{f}{\rm yr^{-1}} \right)^{-2/3}\,. \tag{103}$$

In particular, for a finite number of sources in a comoving volume V_c with the rest frame frequency in the range of $f_{\rm min} < f_{\rm r} < f_{\rm max}$, Equation 102 can be re-written as follows,

$$h_c^2(f) = \frac{4\pi^{-1/3}}{3c^2} f^{-4/3} \sum_{\substack{i \\ f_{\rm min} < f_{\rm r} < f_{\rm max}}} \frac{1}{(1+z_i)^{1/3}} \frac{(G\mathcal{M}_i)^{5/3}}{V_c} \frac{N_{\rm galaxy,total}}{N_{\rm galaxy}} \tag{104}$$

where i represents each GW source (BH merger event) in the galaxies of both mass ranges. Assuming that our galaxy sample of size $N_{\rm galaxy}$ is representative of the properties of the entire set of galaxies in the Millennium simulation of number $N_{\rm galaxy,total}$, we normalize our estimate of the strain with a factor of $N_{\rm galaxy,total}/N_{\rm galaxy}$. The variable $f_{\rm min}$ represents twice the Keplerian orbital frequency calculated with the values of the binary mass and the semimajor axis at the time when GWs become dominant to shrink the orbit (i.e., when the merger condition 1 is fulfilled). For $f_{\rm max}$, we assume the frequency at the innermost circular orbit, or $f_{\rm max} = [2/(1+z)] f_{\rm orb}(r = 3r_{\rm sch})$ (Hughes, 2002; Ravi et al., 2012; McWilliams et

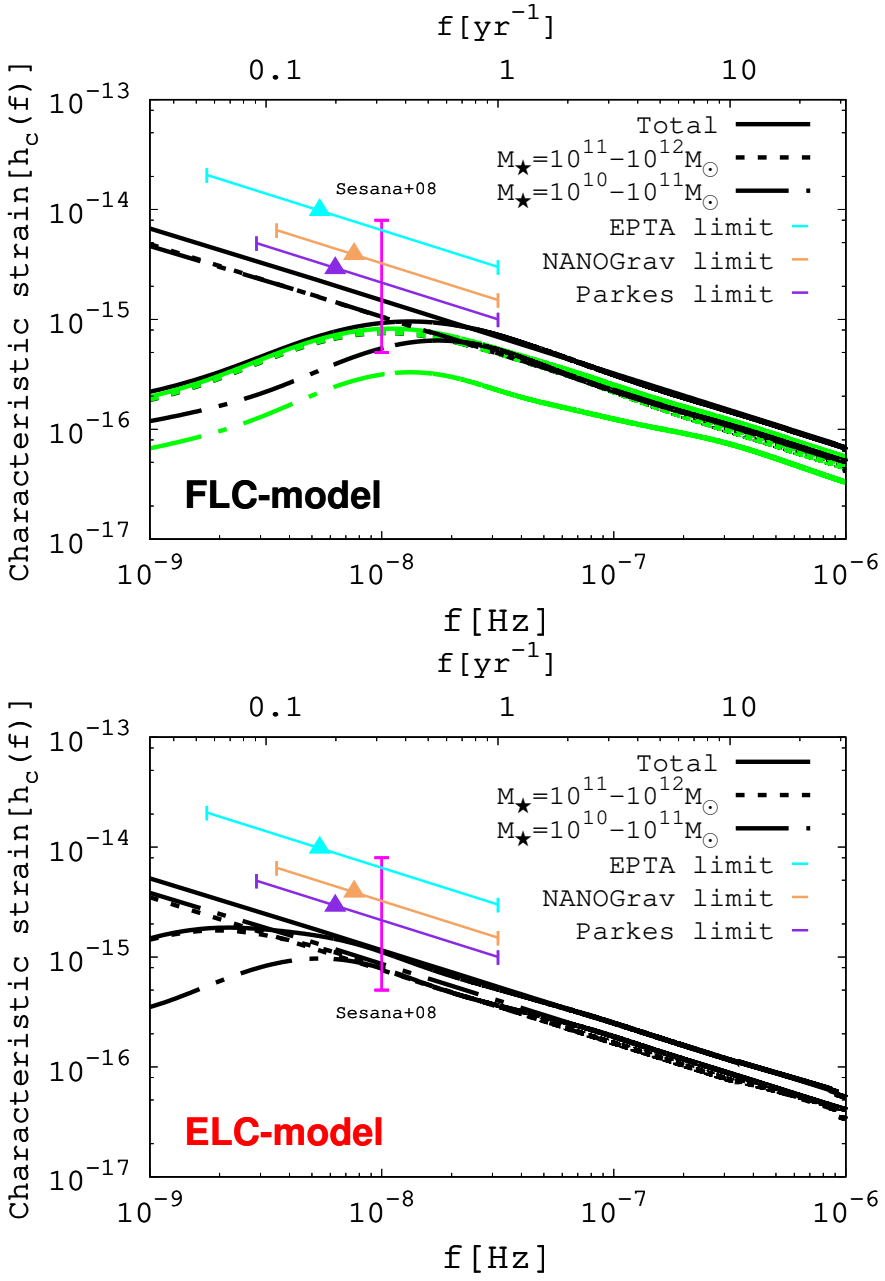

Figure 35: The spectra of the characteristic strain from the galaxies of each mass range (dotted line for $M_\star = 10^{11} - 10^{12}$ $M_\odot$ and dot-dashed line for $M_\star = 10^{10} - 10^{11}$ $M_\odot$) as well as the total estimate (solid line) for the FLC-model (*left* panel) and the ELC-model (*right* panel). The straight lines represent the strain assuming circular orbits, while the curved lines show the modification when the orbital eccentricities are taken into account. As a reference, we also indicate the upper limits for the strain with thin solid lines. The triangles show the upper limit in each experiment at its peak sensitivity. In the *left* panel, the green lines indicate the spectra made with only SMBH binaries fulfilling the bypass condition (Equation 101). As a consequence of three-body interactions, the chirp mass is higher for BH mergers in the ELC-model (see Figure 31). As a result, despite the lower BH merger rates, BH mergers in the smaller galaxies of $M_\star = 10^{10} - 10^{11}$ $M_\odot$ almost equally contribute to the GWB as those in more massive galaxies.

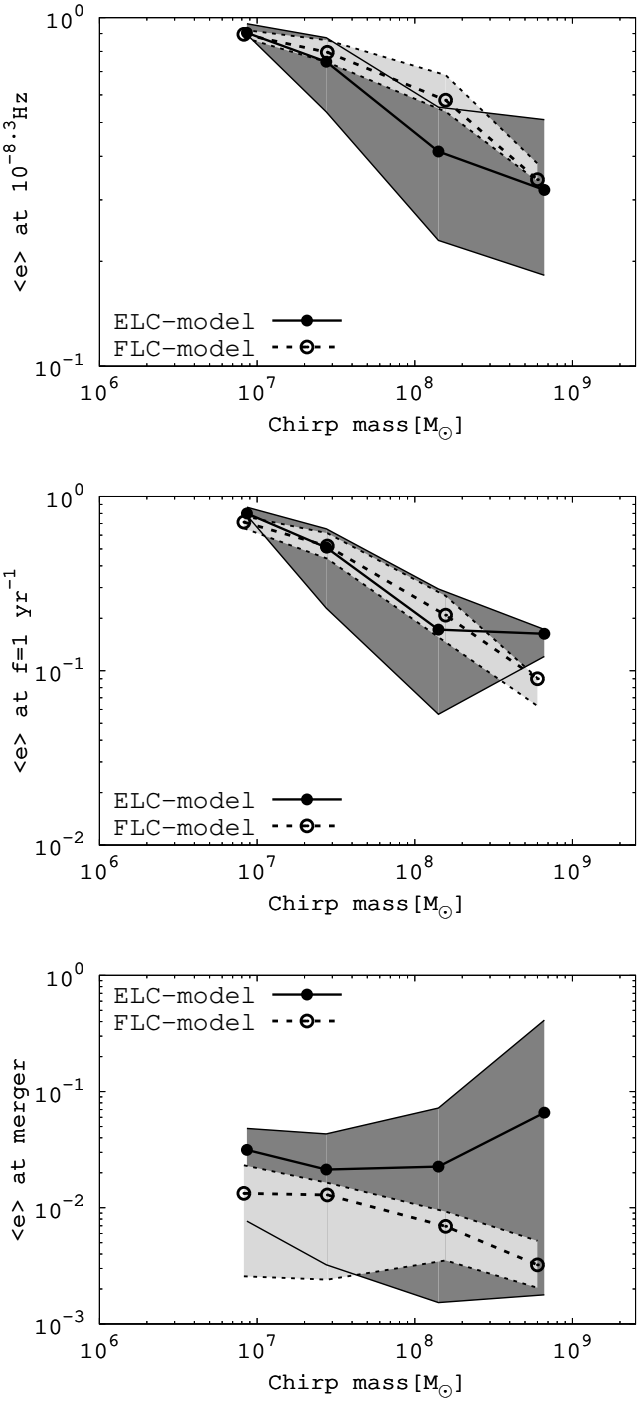

Figure 36: $M_{\rm chirp}^{5/6}$-weighted (corresponding to weighting by GW intensity) average e as a function of $M_{\rm chirp}$ for the FLC-model (dotted line with hollow circles) and the ELC-model (solid line with solid circles) at three characteristic frequencies, i.e. $f = 10^{-8.3}$ Hz (near peak sensitivity, see Figure 37), $f = 1$ yr^{-1} Hz $= 10^{-7.5}$ Hz and f at merger. We analytically estimate the eccentricities at which GW emission becomes more efficient (Peters & Mathews, 1963). The shaded regions indicate 68% of BH mergers at a given chirp mass.

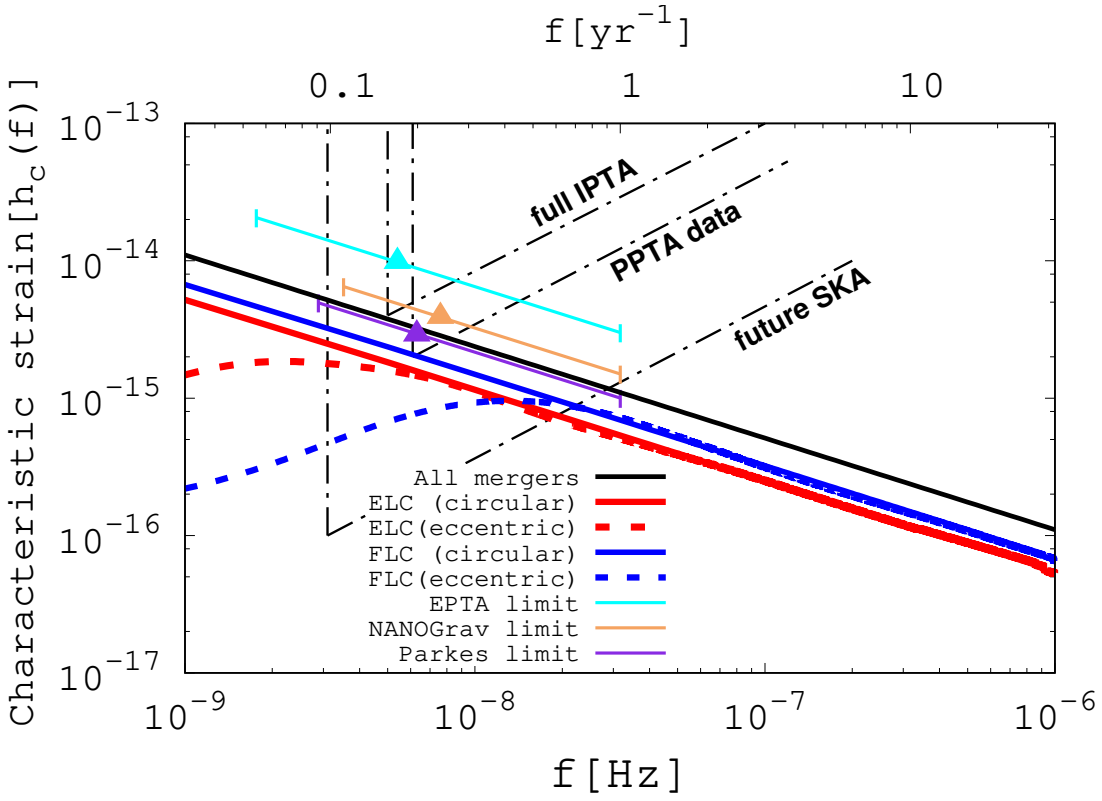

Figure 37: Our estimates for h_c are compared with the current/future-expected sensitivities (wedge-shaped dot-dashed lines). The upper-most wedge indicates the sensitivity set by the full IPTA ($h_c \sim 10^{-16} - 10^{-17}$ at $f = 5 \times 10^{-9}$ Hz, Janssen et al., 2015). The predictions of our models cannot yet be tested with the current instruments. The other two lines refer to the sensitivity set by complete Parkes PTA (PPTA, 20 pulsars for 5 yr) data set (labeled "PPTA data") and that achievable with the planned Square Kilometer Array (SKA) assuming monitoring of 20 pulsars over 10 yr (labeled "future SKA"). The sensitivity provided by the PPTA dataset may not be sufficient to reach the strain inferred from our models. In the future, we expect that the planned SKA will be able to impose constraints over wider frequency ranges.

al., 2014), which is written as,

$$f_{\rm max} = 2.2 \times 10^{-5} \left(\frac{M_{\rm BH,1}}{10^8\ {\rm M}_\odot}\right)^{-1} \left(1 + \frac{M_{\rm BH,2}}{M_{\rm BH,1}}\right)^{1/2} {\rm Hz}\,. \tag{105}$$

Assuming circular orbits and given the amplitude scaling as $f^{-2/3}$ (see Eq.103), from our merger rates we find that $A = 0.70 \times 10^{-15}$ for the FLC-model and $A = 0.53 \times 10^{-15}$ for the ELC-model. We show our estimates for the characteristic strain $h_{\rm c}$ for the FLC-model (denoted by "FLC (circular)") and the ELC-model (denoted by "ELC (circular)") in Figure 34. The black solid line (labeled "All mergers") above the two lines for the FLC-model and the ELC-model corresponds to the strain assuming all galaxy mergers lead to BH mergers given the sampled galaxy merger trees (see more details in §5.7.3). We additionally depict the GW spectra inferred in other published studies (Jaffe & Backer, 2003; Sesana et al., 2008; Kocsis & Sesana, 2011; Shannon et al., 2013; McWilliams et al., 2014; Kulier et al., 2015) and observational upper limits set by EPTA ($A = 3.0 \times 10^{-15}$, Lentati et al., 2015), NANOGrav ($A = 1.5 \times 10^{-15}$, Arzoumanian et al., 2016) and Parkes ($A = 1.0 \times 10^{-15}$, Shannon et al., 2015). For the latter, we indicate the upper limit in each experiment at its peak sensitivity (triangles), and we extrapolate this limit to other frequencies assuming a power law of $f^{-2/3}$. The frequency range shown in each case is $1/T < f < 1\ {\rm yr}^{-1}$, where T is the total observing time. In spite of different dominant mechanisms for orbital decay in the FLC- and ELC- model, the values are comparable. We believe the reasons are as follows:

1. The merger rates for the ELC-model are lower at $0.5 < z < 2$ than those for the FLC-model (see Figure 32). The resulting decrease in the GWB, however, is relatively minor, because it is the mergers involving the lowest-mass BHs that are missing. The mergers which dominate the GWB, involving more massive BHs, are still occurring in the ELC-model.

2. In the ELC-model, we find that binaries have longer lifetimes (see §5.5.2) due to the time taken for multiple SMBHs to accumulate in the cores as the host galaxies go through successive mergers. This can cause an overall delay of the BH mergers compared to nearly prompt mergers in the FLC-model. This results in sparse mergers at higher z and, more importantly, copious GW emissions at lower z. Moreover, individual GW emissions are more powerful because the delay of mergers causes BHs to accrete more mass before they undergo mergers. This would compensate for the decrease in the GWB due to the loss of some BH mergers, as described in 1 above.

3. In the ELC-model, BH mergers in the smaller galaxies of $M_\star = 10^{10} - 10^{11}$ M$_\odot$ contribute more to the GWB than those in more massive galaxies. Those contributions are even higher than those for the FLC-model because of the more frequent mergers with larger chirp mass at lower z in the ELC-model. Figure 35 shows how much BH mergers in the galaxies of each mass range contribute to the total estimates for the FLC-model (*left* panel) and the ELC-model (*right* panel). The dotted line for $M_\star = 10^{11} - 10^{12}$ M$_\odot$ and the dot-dashed line refers to $h_{\rm c}$ for $M_\star = 10^{10} - 10^{11}$ M$_\odot$ assuming circular orbits. As shown in the *right* panel, the strain is higher in the ELC-model from BH mergers in smaller galaxies.

Also note that, as will be explained in §5.7.2, the curved lines show the effect on the strain of binary eccentricities.

We also find that the amplitude of the characteristic strain is dominated by BH mergers at low redshift $z < 2$ (see also Wyithe & Loeb 2003). In the FLC-model, 86% of the BH coalescences occur at $z < 2$ with an average chirp mass of $\mathcal{M} = 1.0 \times 10^8$ $M_\odot$, while in the ELC-model the fraction of mergers at $z < 2$ is 98% with $\mathcal{M} = 1.6 \times 10^8$ $M_\odot$. If we impose more stringent constraints on z, in the FLC-model the fraction decreases to 65% with $\mathcal{M} = 1.2 \times 10^8$ $M_\odot$ at $z < 1$ and to 35% with $\mathcal{M} = 1.4 \times 10^8$ $M_\odot$ for $z < 0.5$. In the ELC-model the fraction becomes 79% with $\mathcal{M} = 1.8 \times 10^8$ $M_\odot$ and 49% with $\mathcal{M} = 2.1 \times 10^8$ $M_\odot$. However, still the majority of SMBH binaries effectively emit GWs at $z < 1$. The increase in the chirp mass especially for the ELC-model can be seen in the redshift evolution of the average chirp mass shown in Figure 31. Here, we separately show the results for the galaxies of each mass range, but the average chirp mass (including the fraction of BH mergers) given above is estimated based on all merger events for galaxies in both mass bins.

5.7.2 Stochastic GW background - GWs from eccentric orbits

An eccentric orbit emits GWs at all integer harmonics of the orbital frequency (Peters & Mathews, 1963; Peters, 1964). Especially for very eccentric orbits, the GW radiation power is greater at higher harmonics. Since the evolution of a binary orbit strongly depends on the evolution of the eccentricity, this may change the shape of the spectrum. In fact, larger contributions from higher harmonics effectively suppress power at lower frequencies, leading to a low frequency flattening or even a turnover in the spectrum (Enoki & Nagashima, 2007; Sesana, 2010, 2015). Therefore, it is necessary to take into account such effects of the eccentricity for more realistic estimates of the GWB.

We find that the binary orbits are very eccentric when GW emission becomes more efficient in our simulations (see Figure 25). In Figure 36, we show $M_{\rm chirp}^{5/6}$-weighted average e as a function of $M_{\rm chirp}$ for the FLC- and ELC-model at three characteristic frequencies: $f = 10^{-8.3}$ Hz (near peak sensitivity), $f = 1$ yr$^{-1} = 10^{-7.5}$ Hz and f at merger. Given that these are eccentricities at which GW emission plays a dominant role in causing the decay of the binary orbits, we analytically estimate the eccentricities at those frequencies (Peters & Mathews, 1963). The shaded regions indicate 68% of BH mergers at a given chirp mass. Generally speaking, the eccentricities for the ELC-model tend to have large scatter compared to those for the FLC-model. The eccentricities are still quite high at $f = 10^{-8.3}$ Hz and $f = 1$ yr^{-1}.

To account for such high eccentricities, we have to consider harmonics up to $n_{\rm max} \simeq 10(1-e)^{-3/2}$ (for $e = 0.99$, $n_{\rm max} \simeq 10000$), which means that a direct summation of the contributions of each harmonic is computationally expensive. We instead emply the fitting formula (16) given in Chen et al. (2017), which has been shown to successfully reproduce the spectrum within a maximum error of 1.5% in log amplitude (i.e. 3.5% in amplitude) for a reference case ($e = 0.9$). The thick dotted lines in Figure 34 and Figure 37 show the spectra when high eccentricities are taken into account. As expected, the spectra at lower frequency of $f < 1$ yr^{-1} are flattened and turn over towards lower f. The strain for both models predicted under the assumption of circular orbit is hardly distinguishable. However, clear

deviations between the two models can be seen when the different eccentricity evolutions are taken into account in the computation. Eccentric spectra start differing from their circular counterparts at frequencies of $f \sim 10^{-7.5}$ Hz ~ 1 yr^{-1} in both models, and display maxima in the region around $f < 10^{-8}$ Hz. Such turnovers of the spectra are consistent with the spectra predicted for $e = 0.9$ in Enoki & Nagashima (2007). The overall shapes of the spectra are also similar with what is found for the case of initially very eccentric binaries in a dense stellar environment (see Figure 3 in Sesana 2015). The spectra from BH mergers in galaxies of both mass ranges are comparable at frequencies of $f > 10^{-8}$ Hz; however at $f < 10^{-8}$ Hz, the signals from more massive galaxies are clearly larger.

We also have checked how the spectra are altered when we exclude SMBH binaries not fulfilling the bypass condition (Equation 101). This exclusion rules out 3 binaries from each galaxy mass bin. Interestingly, we find that $A = 0.54 \times 10^{-15}$ at $f = 1$ yr^{-1}, which is even closer to that for the ELC-model. The green lines in the *left* panel of Figure 35 show the modified spectra as a result of the exclusion. The turnovers are now less pronounced and shifted to lower frequencies of $f < 10^{-8}$ Hz.

As PTA observation periods span decades, the frequency range of $f \sim 10^{-9} - 10^{-8}$Hz is most sensitive to GWs. In Figure 37, the current and future-expected sensitivities and observational upper limits are compared with our estimates. The uppermost edge indicates the sensitivity by the full IPTA ($h_c \sim 10^{-16} - 10^{-17}$ at $f = 5 \times 10^{-9}$ Hz, Janssen et al., 2015). The other two lines refer to the sensitivity set by the complete Parkes PTA (PPTA) data set with 20 pulsars for 5 yr and that achievable with the planned Square Kilometer Array (SKA) with 20 pulsars over 10 yr [27]. *Our models predict amplitudes below the observational upper limits. The current PPTA dataset may not be sufficient to confirm/rule out our models; however in the future, the planned SKA will be able to give constraints on our models over wider frequency ranges.*

5.7.3 Semi-analytic analysis on the estimate of A - Comparison with previous works

In this work, using few-body simulations in analytic background potentials, we follow the dynamical evolution of multiple SMBH systems and estimate the BH coalescence rates in the host galaxies undergoing multiple mergers with a wide range of mass ratios. Using the computed merger rates, we next estimated the stochastic GWB. For a more thorough understanding of our results, it is hence important to compare our results with previous works.

As an informative comparison, given our sampled merger trees, we analytically estimate A following the assumptions about BH mergers in McWilliams et al. (2014) and Kulier et al. (2015). They assume that,

1. *every bound pair of BHs efficiently solves the final parsec problem on its own*;
2. *BH binaries are always in circular orbits.*

[27] The expected level to be reached by the SKA is lower, $h_c \sim 10^{-16} - 10^{-17}$ at a reference frequency of yr^{-1} (Janssen et al., 2015).

While these calculations broke new ground in estimating BH merger rates in a cosmological context, their assumption 1 is optimistic, and their predicted GW emission rates should be considered upper limits. As a result, the GWB predicted by McWilliams et al. (2014) and Kulier et al. (2015) is higher than the one given by our detailed computations. By comparing A for the optimistic case to A for the FLC and the ELC models, we may be able to understand how much each assumption affects A. For this estimate, we additionally assume that BH mergers occur after a dynamical friction timescale (equation 3 in Kulier et al. 2015) since galaxy merger events. This leaves 17 out of our total 1744 mergers incomplete by $z = 0$. We take the total mass of merging binaries to be the maximum value between the BH mass required by the $M - \sigma$ relation at the BH merger redshift and the sum of the masses of two merging BHs. We find the total $A = 1.10 \times 10^{-15}$ for the optimistic case, which is larger by a factor of $1.5 - 2$ than those for the FLC-model and the ELC-model. The total A decreases as we impose $q_\star$ cutoffs: Assuming only galaxy mergers of $q_\star > 0.01$ lead to BH mergers, $A = 1.00 \times 10^{-15}$. For $q_\star > 0.1$, $A = 0.77 \times 10^{-15}$, which is fairly close to A for the FLC-model. As shown in Figures 26 and 32 (galaxy merger rates higher than BH merger rates, see (iii) in §5.6.2), we can confirm again that not all galaxy mergers lead to BH mergers in both the FLC and the ELC models, resulting in smaller A. In Figure 38, we show the evolution of A with redshift for the optimistic case (labeled "all mergers"), the FLC- and the ELC-model. In addition to the total A (*upper* panel), we separately show the evolution of A for more massive galaxies (*middle* panel) and less massive galaxies (*bottom* panel). In all three panels, we can see that the amplitudes for the FLC-model and the ELC-model remain below those for the optimistic case. Due to nearly prompt BH mergers for the optimistic case and delayed mergers in the ELC-model, the first GW signals for the optimistic case appear earliest, followed by those for the FLC-model and the FLC-model at last. The first mergers in the ELC-model are delayed by $\Delta z \simeq 0.3 - 1.5$ with respect to those in the FLC-model.

If we relax the assumption of circular orbits, as seen in Figures 34 and 35, the GWB further decreases, especially at low frequencies. In our two models, the effects of the eccentricities at $f = 1$ yr^{-1} are not significant, but the difference could exceed an order of magnitude at lower frequencies depending on whether merged binary orbits are assumed to be circular or eccentric.

A suppression of the GW signal at higher frequencies can be caused by the presence of a circumbinary disk. In another recent study, Kelley et al. (2017), using the galaxy population in the Illustris simulation, coevolve massive BHs to predict the GWB. They take into account various possible environmental mechanisms in their calculations including dynamical friction, stellar 'loss cone' scattering and tidal-viscous drag from a circumbinary disc. Similarly to our models, they explore different degrees of loss cone filling. Their fiducial model predicts an amplitude within the range of $0.3 \times 10^{-15} < A < 0.4 \times 10^{-15}$ (with the upper limit of $A \simeq 0.7 \times 10^{-15}$). This is smaller than our values roughly by a factor of $1 - 2$. We believe that this may be caused by different strategies to populate SMBHs (Sesana et al., 2016). Furthermore, tidal torques from the gaseous circumbinary disk could also come into play. This was studied before by Kocsis & Sesana (2011) with BH merger rates from the Millennium simulation and adopting the models for gas-driven inspirals of Haiman et al. (2009). Typically, the presence of circumbinary discs drives very rapid inspirals of binaries via migration, leading to a significant suppression of the signal at frequencies $f > 10^{-8}$ Hz compared to mergers in a gas-poor environment.

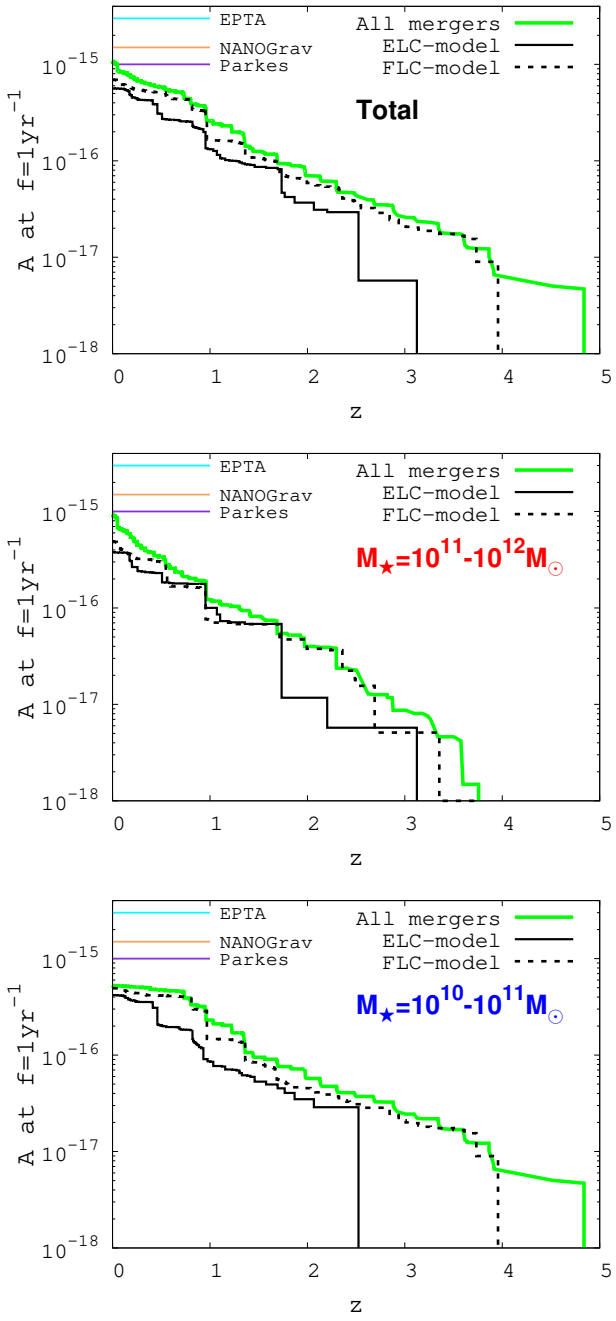

Figure 38: The redshift evolution of the GWB amplitude A for the optimistic case (labeled "all mergers", thick green solid line), the FLC-model (dotted black line) and the ELC-model (thin solid black line). The *top* panel shows the total A and we separately depict the evolution of A contributed by more massive galaxies (*middle* panel) and less massive galaxies (*bottom* panel). As a reference, the observational upper limits for EPTA, NANOGrav and Parkes are indicated.

Generally speaking, adopting the scaling relations to populate SMBHs in the galaxies, our estimates for A are generally in good agreement with other studies (see Figure 34), especially with models constructed on the Millennium simulation (e.g Sesana et al., 2009). However, noting the discussion in Sesana et al. (2016) and Rasskazov & Merritt (2017), we emphasize that our results could also be affected by the use of different observational relations. Sesana et al. (2016) employ different SMBH-galaxy scaling relations and accretion prescriptions to populate and grow the SMBHs, and they study the impact of selection bias in determining SMBH masses on the PTA measurements. In another study, Rasskazov & Merritt (2017), taking into account the effects of rotating and aspherical nuclei on the orbital evolution of SMBH binaries, compute the GWB and study the dependence on the $M_{\rm BH} - M_\star$ relation. Even though they tackle this problem within different frameworks, both studies suggest that the GWB amplitude has been overestimated and may decrease by a factor of a few if different galaxy scaling relations are used [28].

After the original submission of this manuscript, we became aware of a similar recent study by Bonetti et al. (2017). They employ a semi-analytic model of galaxy evolution and model SMBH mergers and their GW signals, by incorporating three-body PN effects to study the role of triple and quadruple interactions between SMBHs (adopted from Bonetti et al. 2016). Their inferred merger rates are fairly consistent with those shown in §5.6, and the physical picture of the mergers we find in our work (see §5.5.1) is also similar to that discussed in Bonetti et al. (2017).

5.7.4 Caveats

Our results were obtained in models with observationally and numerically-motivated assumptions, but they are subject to several caveats. In this section, we discuss the major limitations of our models.

1. *Dynamical friction.* In the FLC-model, we assume dynamical friction operates very efficiently to decay the orbits of BH binaries at all times. This regime may underestimate the true hardening rate in the presence of a full loss cone, since inside the influence radius stellar scattering hardens SMBH binary orbits by a factor $\sim 1/q$ faster than the hardening rate from a naive application of the dynamical friction formula (Merritt, 2013). Furthermore, our merging galaxy model assumes a flat core in the inner region, not accounting for the dynamical changes in shape of the inner stellar potential as SMBH binaries in it mutually evolve. The shape of the stellar potential around SMBH binaries is correlated with the hardening rates of the binaries (Dosopoulou & Antonini, 2017), hence their fate and the BH merger rates in the FLC-model.

2. *Three-body interactions*

   We have not self-consistently included the 3-body PN terms (see Equation 5 in Lousto & Nakano 2008) during the 3-body interactions (e.g Bonetti et al., 2016). However, as explained in §5.5.1, most of the important 3-body interactions are in the hierarchi-

[28] Taylor et al. (2016) discuss the similar issue of the overestimate of SMBH merger rates from an observational perspective.

cal regime, with only 2 of the 3 bodies sufficiently close for PN terms to be needed. Therefore we expect that our results are robust, but we will explore this in future work.

3. *Assumptions on SMBH mass growth*

   There are several factors related to the assumptions on SMBH mass growth and sampled galaxies which may affect the GWB. First of all, given the requirement for the total mass of the central BHs, very loud signals from a few massive binary mergers of $q \simeq 1$ at $z \simeq 0$ can cause remarkably large jumps in the GWB. As explained in §5.3.2, this could occur when binaries, with the central BHs and other initially small BHs which rapidly grow in the cores while the central BHs are temporarily ejected, coalesce. Such temporary ejection of the central BHs can occur as a result of GW recoil kicks. Given our galaxy samples, we do not find that A at $z = 0$ is dominated by a few of these loud GW emission events. But it is still possible at lower z, especially more likely in the ELC-model with its more frequent ejections[29]. If we explore a large number of galaxy merger trees, the statistical errors from finite sample size[30] will decrease, but chances of large signals from a few individual sources may increase. Furthermore, in this study we do not consider large kicks driven by nonlinear spin-orbit interactions (Lousto & Zlochower, 2011; Lousto et al., 2012; Lousto & Zlochower, 2013) with the maximum recoil velocities larger than typical escape velocities of galaxies. We point out that the frequencies of ejection and merger events would vary when such large kicks are taken into account.

4. *Galaxy samples*

   We have not included the contributions from smaller galaxies of $M_\star < 10^{10}$ $M_\odot$ (or $M_{\rm DM} < 5 \times 10^{11}$ $M_\odot$). We find that BH mergers from less massive galaxies contribute more to the overall GWB as a result of higher BH merger rates as shown in Figure 33. Therefore, it is also possible that BH coalescences in galaxies with $M_\star < 10^{10}$ $M_\odot$ can noticeably increase the predicted GWB. However, it is uncertain whether the BH merger rates increase further for galaxies of $M_\star < 10^{10}$ $M_\odot$ and, more importantly, the offset between the increase in the BH merger rates and the decrease in the chirp mass lead to a significant rise in the GWB. For such low mass galaxies, the relationship between stellar mass $M_\star$ and halo mass $M_{\rm DM}$ is more complicated than the single power law we have assumed(e.g Behroozi et al., 2010; Moster et al., 2013), and our model would need to be modified accordingly. In addition, we assume that each galaxy always harbors a SMBH as long as the BH mass is larger than the minimum mass. However, the occupation fraction in the low mass galaxies is more likely to be affected by the assumptions on BH seed formation and initial occupation fraction at high redshift (Menou et al., 2001). Therefore, considering that less massive halos tend to possess relatively small number of early progenitors as well as even small mass progenitors,

[29]In the FLC-model, we find that the BH merger mass ratios decrease and the BH masses grow as z decreases, following the trend in the galaxy merger histories. Hence GW kick-driven ejections, with subsequent rapid growth of small BHs, are more likely at higher z. However, their contributions may not be significant to A at $z = 0$ and chances of such giant binary formations and mergers at small z may be low.

[30]The Poisson error of $\sqrt{N_{\rm BH\ merger}}/N_{\rm BH\ merger} \simeq 0.1$, where $N_{\rm BH\ merger}$ is the number of BH mergers for the current sample size.

the contributions to the GWB from low mass galaxies of $M_\star < 10^{10}$ $M_\odot$ may not be significant (Sesana, 2013). But these estimates are approximate, so more systematic studies are necessary for better understanding of the contributions of the BH mergers in dwarf galaxies.

Another caveat is that, given that we follow up to $10 - 12$ galaxy mergers, for host galaxies experiencing a large number of mergers at low redshift, we may miss some galaxy mergers in their histories, hence possibly leading to an underestimate of the GWB.

Last, in our merging galaxy model, we assume one central BH per satellite galaxy at galaxy mergers. Multiple BHs in satellite galaxies are definitely possible. For those cases, more frequent multi-BH interactions and ejections will take place, which possibly influence the BH merger rates, ejection rates as well as the chances of such giant binary mergers explained above.

5.7.5 Summary

In this work, using few-body simulations in analytic background potentials, we have examined the evolution of SMBH binaries and higher multiples, from their formation to coalescence, as the host galaxies go through mergers with mass ratios of $10^{-4} < q_\star < 1$. For host galaxies of mass $M_\star = 10^{10} - 10^{12}$ $M_\odot$ at $z = 0$ extracted from the Millennium simulation, we followed their merger trees by assuming a SMBH in each of the host/satellite galaxies, with the BH mass determined by standard scaling relations. We have explored two limiting scenarios for the decay of the binary orbits, approximating full and empty loss cone regimes. In the full loss cone model (FLC-model), we assume dynamical friction efficiently shrinks the orbits until binaries merge, whereas in the empty loss cone model (ELC-model), we assume that dynamical friction is no longer able to operate and cause orbital decay when the binaries become hard. The ELC-model allows us to examine multi-body BH interactions in a cosmological context, and test their utility as a "solution of last resort" to the final parsec problem in large galaxies where other solutions may fail. The FLC-model, while highly idealized, serves as a valuable comparison case, and as a testing ground for an underexplored regime: inspirals where e is excited to very high values by dynamical friction in a flat stellar core characteristic of the highest mass galaxies. Based on the inferred merger rates from our simulations, we estimate the stochastic GWB in the two models. We summarize our findings as follows:

1. *Dynamical features of SMBH binaries and multiple systems*: we find a few clear differences in the evolution of SMBH binaries between the FLC-model and the ELC-model. For the FLC-model, dynamical friction tends to increase the binary eccentricity. When energy loss to GWs becomes dominant, the binary eccentricities are almost unity ($e > 0.99$). The evolution of the orbital eccentricity of SMBH binaries in various stellar distributions will be explored in a future paper (Stone et al. 2017, in prep). However, our FLC results can be understood in the context of past work, which finds eccentricity excitation due to dynamical friction in flat density profiles, particularly in Keplerian potentials (Dosopoulou & Antonini, 2017). A critical assumption behind the eccentricity excitation seen in the FLC model is the existence of a flat stellar density

core. While this assumption is reasonable for the very high-mass galaxies considered in this paper, it would not apply to lower-mass galaxy mergers. Subsequently, SMBH binaries merge on a short time scale and the lifetimes of coalescing binaries are less than 1 Gyr. We also find that the BH merger events are strongly coupled with major mergers ($q_\star > 0.1$) of the host galaxies (Figure 26). For the ELC-model, however, there is a time delay before the central SMBH binaries merge because they must wait for other BHs to come close and effectively interact with them. This results in longer binary lifetimes ($\gtrsim$ 1 Gyr). This is a clear difference from the FLC-model. As a result, coalescences of SMBHs in the ELC-model preferentially occur in the host galaxies experiencing multiple major mergers.

2. *BH merger rates*: we find that SMBH binaries do merge in both models, but with typically higher coalescence rates in the FLC-model than in the ELC-model. *There is no "final parsec" problem in either scenario.* Even though the BH coalescence rate for the ELC-model is lower, the BH mergers in the ELC-model strongly indicate that, as galaxies go through a series of mergers and binaries stall due to inefficient decay mechanisms (e.g., empty loss cone), a cluster of multiple SMBHs is naturally produced in the core regions, and these BHs can eventually merge via multi-body interactions.

3. *Mass ratio of coalesced BH binaries*: another noticeable feature of the ELC-model is that the mass ratios, and hence the chirp masses, of coalescing SMBHs tend to be higher. As they go through chaotic three-body interactions, the less massive BHs will typically be ejected, leaving behind a binary of the more massive BHs.

4. *Stochastic GW background*: using the inferred BH coalescence rates, we estimate the strain amplitudes $A = 0.70 \times 10^{-15}$ and $A = 0.53 \times 10^{-15}$ for the FLC-model and the ELC-model, respectively. In spite of the lower BH merger rates for the ELC-model, we find that the amplitudes are quite similar. This is because (a) mergers of BH binaries, especially with large chirp masses, still occur in the ELC-model. Only those with lower mass ratios, which make minor contributions to the GWB, are missing; (b) in the ELC-model, BH coalescence events preferentially take place at a later time with larger chirp mass, as BHs have more time to grow, given the delayed mergers. In other words, louder GW emissions more abundantly occur at lower redshift. This would counterbalance the decrease in the GWB due to the loss of some BH mergers, as described in 1 above; (c) due to the larger mass ratios of the merged binaries, the contributions of the less massive binary mergers (i.e., coalesced BHs in less massive galaxies) to the GWB are relatively high in the ELC-model compared to the FLC one. Our inferred strain is consistent with current observational limits and a factor of roughly two below the rates predicted by the simple model in which every galaxy merger leads to a BH merger.

5. *Effect of high eccentricities on GW spectra*: Given the high eccentricities of the merged SMBH binaries, our models predict significant suppression of GW power at lower frequencies. This causes a low frequency flattening as well as a turnover in the stochastic background spectrum as shown in Figures 34 and 35, which will be observationally important for comparison to future data.

By adopting a dynamical approach to study the coalescence of SMBH binaries, our work shows clear distinctions between two limiting regimes of loss cone physics. In particular, different expectations for chirp masses, mass ratio distributions, and flattening of the GW spectra due to high eccentricities can all be observationally relevant. Furthermore, multi-body interactions between SMBHs are a natural consequence of galaxy mergers, and are clearly a plausible channel for driving BH coalescences. Our predictions show that ongoing PTA searches can potentially discriminate between different models of black hole binary orbital evolution.

6 Conclusions

Black hole binaries are excellent astrophysical laboratories that will provide a fundamental understanding of black hole physics. Furthermore, they are an important ingredient to interpret the history of the Universe. In this thesis, we have studied the formation of BH binaries at different cosmic times since the first stars collapsed to a BH, focusing on their interactions with surroundings and their global-scale observational signatures across the history of the Universe. A broader overview of them is provided in §1.

We first present a multi-purpose N-body code in §2 developed for the studies outlined below. The code can solve the equations of motion with physical effects including mutual gravitational force up to the 2.5-order post-Newtonian terms, dynamical friction, force from a background potential and mass growth. Furthermore, by implicitly estimating the energy variations from various dissipative mechanisms, the code can account for the dynamical evolution of any integrable background potentials. We employ this code for performing standard scattering experiments and studying the formation of runaway stars.

In §3, we investigate the formation of Population III X-ray binaries and their X-ray output. We estimate their formation rate, using N-body simulations, which is the first estimate of this kind. We show that based on the estimated formation rate, the X-ray luminosity of HMXBs per star formation rate is much higher than what has been observed in the local Universe. This implies that X-rays can heat the IGM rapidly at Cosmic Dawn, which can be probed via 21 cm line detection.

In §4, we study the formation of IMBHs and SMBHs via "hyper-Eddington" accretion in centrally condensed gas clouds from Population III remnant BHs. We find that only one 10^3-10^5 $M_\odot$ IMBH can form via such rapid accretion per galaxy, and that this IMBH typically captures a stellar-mass BH companion. This can be a viable pathway to the formation of EMRIs, which are important gravitational wave sources of eLISA.

Finally, in §5, we examine the evolution of SMBH binaries and higher multiples, from their formation to coalescence. We show that BHs merge via dynamical friction and multi-body interactions between SMBHs. In this study, we find that multi-BH interactions could be a promising solution to the final parsec problem. Furthermore, we discuss the stochastic gravitational wave background from their mergers. In particular, we compute the effects of high eccentricities on the GWB spectrum.

Over the last decades, great scientific advance in observation and theory as well as numerical techniques have been made. Those have added new aspects to our understanding. And every new observational and theoretical results will continue to provoke new ideas for

solving many astrophysical problems as they have been doing so far. In conducting the studies presented in the dissertation, we have frequently encountered many interesting questions. These will be explored with new ideas and new exciting results from a number of ongoing as well as planned surveys for observing the high-z Universe and gravitational wave emission from BH mergers.

References

Abel, T., Bryan, G. L., & Norman, M. L. (2000). *ApJ*, *540*, 39-44.

Abel, T., Bryan, G. L., & Norman, M. L. (2002). *Science*, *295*, 93-98.

Abramowicz, M. A., Czerny, B., Lasota, J. P., & Szuszkiewicz, E. (1988). *ApJ*, *332*, 646-658.

Agarwal, B., Khochfar, S., Johnson, J. L., Neistein, E., Dalla Vecchia, C., & Livio, M. (2012). *M.N.R.A.S.*, *425*, 2854-2871.

Alexander, T., & Natarajan, P. (2014). *Science*, *345*, 1330-1333.

Alvarez, M. A., Wise, J. H., & Abel, T. (2009). *ApJL*, *701*, L133-L137.

Amaro-Seoane, P., Aoudia, S., Babak, S., Binétruy, P., Berti, E., Bohé, A., ... Wardell, B. (2013). *GW Notes, Vol. 6, p. 4-110*, *6*, 4-110.

Antonini, F., & Merritt, D. (2012). *ApJ*, *745*, 83.

Armitage, P. J., & Natarajan, P. (2002). *ApJL*, *567*, L9-L12.

Artale, M. C., Tissera, P. B., & Pellizza, L. J. (2015). *M.N.R.A.S.*, *448*, 3071-3080.

Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S. J., Chatterjee, S., Christy, B., ... NANOGrav Collaboration (2016). *ApJ*, *821*, 13.

Bañados, E., Venemans, B. P., Decarli, R., Farina, E. P., Mazzucchelli, C., Walter, F., ... Yang, Q. (2016). *ApJ Supp.*, *227*, 11.

Bañados, E., Venemans, B. P., Mazzucchelli, C., Farina, E. P., Walter, F., Wang, F., ... Winters, J. M. (2018). *Nat.*, *553*, 473-476.

Bañados, E., Venemans, B. P., Morganson, E., Decarli, R., Walter, F., Chambers, K. C., ... Waters, C. (2014). *AJ*, *148*, 14.

Bansal, K., Taylor, G. B., Peck, A. B., Zavala, R. T., & Romani, R. W. (2017). *ApJ*, *843*, 14.

Basu-Zych, A. R., Lehmer, B. D., Hornschemeier, A. E., Bouwens, R. J., Fragos, T., Oesch, P. A., ... Zezas, A. (2013). *ApJ*, *762*, 45.

Becerra, F., Greif, T. H., Springel, V., & Hernquist, L. E. (2015). *M.N.R.A.S.*, *446*, 2380-2393.

Begelman, M. C. (1979). *M.N.R.A.S.*, *187*, 237-251.

Begelman, M. C. (2012). *M.N.R.A.S.*, *420*, 2912-2923.

Begelman, M. C., Blandford, R. D., & Rees, M. J. (1980). *Nat.*, *287*, 307-309.

Begelman, M. C., Volonteri, M., & Rees, M. J. (2006). *M.N.R.A.S.*, *370*, 289-298.

Behroozi, P. S., Conroy, C., & Wechsler, R. H. (2010). *ApJ*, *717*, 379-403.

Bekenstein, J. D. (1973). *ApJ*, *183*, 657-664.

Belczynski, K., Kalogera, V., Rasio, F. A., Taam, R. E., Zezas, A., Bulik, T., ... Ivanova, N. (2008). *ApJ Supp.*, *174*, 223-260.

Bellovary, J. M., Governato, F., Quinn, T. R., Wadsley, J., Shen, S., & Volonteri, M. (2010). *ApJL*, *721*, L148-L152.

Benson, A. J. (2005). *M.N.R.A.S.*, *358*, 551-562.

Benson, A. J., & Madau, P. (2003). *M.N.R.A.S.*, *344*, 835-846.

Berentzen, I., Preto, M., Berczik, P., Merritt, D., & Spurzem, R. (2009). *ApJ*, *695*, 455-468.

Binney, J., & Tremaine, S. (1987). *Galactic dynamics*. Princeton University Press.

Blaauw, A. (1961). *Bull. Astron. Inst. Netherlands*, *15*, 265.

Blaes, O., Lee, M. H., & Socrates, A. (2002). *ApJ*, *578*, 775-786.

Blecha, L., Cox, T. J., Loeb, A., & Hernquist, L. (2011). *M.N.R.A.S.*, *412*, 2154-2182.

Bogdanović, T., Reynolds, C. S., & Miller, M. C. (2007). *ApJL*, *661*, L147-L150.

Bonetti, M., Haardt, F., Sesana, A., & Barausse, E. (2016). *M.N.R.A.S.*, *461*, 4419-4434.

Bonetti, M., Sesana, A., Barausse, E., & Haardt, F. (2017). *submitted*.

Boroson, T. A., & Lauer, T. R. (2009). *Nat.*, *458*, 53-55.

Bortolas, E., Gualandris, A., Dotti, M., Spera, M., & Mapelli, M. (2016). *M.N.R.A.S.*, *461*, 1023-1031.

Bowman, J. D., Rogers, A. E. E., & Hewitt, J. N. (2008). *ApJ*, *676*, 1-9.

Bromm, V. (2013). *Reports on Progress in Physics*, *76*(11), 112901.

Bromm, V., Coppi, P. S., & Larson, R. B. (2002). *ApJ*, *564*, 23-51.

Bromm, V., & Loeb, A. (2003). *ApJ*, *596*, 34-46.

Bromm, V., & Loeb, A. (2006). *ApJ*, *642*, 382-388.

Bromm, V., & Yoshida, N. (2011). *Ann. Rev. A&A*, *49*, 373-407.

Brügmann, B., González, J. A., Hannam, M., Husa, S., & Sperhake, U. (2008). *prd*, *77*(12), 124047.

Burns, J. O., Lazio, J., Bale, S., Bowman, J., Bradley, R., Carilli, C., ... Pritchard, J. (2012). *Advances in Space Research*, *49*, 433-450.

Campanelli, M., Lousto, C., Zlochower, Y., & Merritt, D. (2007a). *ApJL*, *659*, L5-L8.

Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D. (2007b). *Physical Review Letters*, *98*(23), 231102.

Cantiello, M., Yoon, S.-C., Langer, N., & Livio, M. (2007). *A&A*, *465*, L29-L33.

Capelo, P. R., Volonteri, M., Dotti, M., Bellovary, J. M., Mayer, L., & Governato, F. (2015). *M.N.R.A.S.*, *447*, 2123-2143.

Chapon, D., Mayer, L., & Teyssier, R. (2013). *M.N.R.A.S.*, *429*, 3114-3122.

Chen, S., Sesana, A., & Del Pozzo, W. (2017). *M.N.R.A.S.*, *470*, 1738-1749.

Chornock, R., Bloom, J. S., Cenko, S. B., Filippenko, A. V., Silverman, J. M., Hicks, M. D., ... Wolfe, A. M. (2010). *ApJL*, *709*, L39-L43.

Clark, P. C., Glover, S. C. O., Smith, R. J., Greif, T. H., Klessen, R. S., & Bromm, V. (2011). *Science*, *331*, 1040-.

Comerford, J. M., Schluns, K., Greene, J. E., & Cool, R. J. (2013). *ApJ*, *777*, 64.

Consortium, T. e., :, Seoane, P. A., Aoudia, S., Audley, H., Auger, G., ... Zweifel, P. (2013). *ArXiv e-prints*.

Cuadra, J., Armitage, P. J., Alexander, R. D., & Begelman, M. C. (2009). *M.N.R.A.S.*, *393*, 1423-1432.

Damour, T., & Deruelle, N. (1981). *Physics Letters A*, *87*, 81-84.

Davis, P. J., Siess, L., & Deschamps, R. (2013). *A&A*, *556*, A4.

de Souza, R. S., Mesinger, A., Ferrara, A., Haiman, Z., Perna, R., & Yoshida, N. (2013). *M.N.R.A.S.*, *432*, 3218-3227.

Deane, R. P., Paragi, Z., Jarvis, M. J., Coriat, M., Bernardi, G., Fender, R. P., ... Rumsey, C. (2014). *Nat.*, *511*, 57-60.

Decarli, R., Dotti, M., Fumagalli, M., Tsalmantza, P., Montuori, C., Lusso, E., ... Prochaska, J. X. (2013). *M.N.R.A.S.*, *433*, 1492-1504.

Desvignes, G., Caballero, R. N., Lentati, L., Verbiest, J. P. W., Champion, D. J., Stappers, B. W., ... Vecchio, A. (2016). *M.N.R.A.S.*, *458*, 3341-3380.

Dijkstra, M., Ferrara, A., & Mesinger, A. (2014). *M.N.R.A.S.*, *442*, 2036-2047.

Dijkstra, M., Gilfanov, M., Loeb, A., & Sunyaev, R. (2012). *M.N.R.A.S.*, *421*, 213-223.

Dijkstra, M., Haiman, Z., Mesinger, A., & Wyithe, J. S. B. (2008). *M.N.R.A.S.*, *391*, 1961-1972.

Dosopoulou, F., & Antonini, F. (2017). *ApJ*, *840*, 31.

Dullo, B. T., & Graham, A. W. (2014). *M.N.R.A.S.*, *444*, 2700-2722.

Dvorkin, I., & Barausse, E. (2017). *M.N.R.A.S.*, *470*, 4547-4556.

Ebisuzaki, T., Makino, J., & Okumura, S. K. (1991). *Nat.*, *354*, 212-214.

Eggleton, P. P. (1983). *ApJ*, *268*, 368.

Emsellem, E., Cappellari, M., Krajnović, D., Alatalo, K., Blitz, L., Bois, M., ... Young, L. M. (2011). *M.N.R.A.S.*, *414*, 888-912.

Enoki, M., & Nagashima, M. (2007). *Progress of Theoretical Physics*, *117*, 241-256.

Erwin, F. (1969). *NASA Technical Report*, *315*, 1-49.

Escala, A., Larson, R. B., Coppi, P. S., & Mardones, D. (2004). *ApJ*, *607*, 765-777.

Faber, S. M., Tremaine, S., Ajhar, E. A., Byun, Y.-I., Dressler, A., Gebhardt, K., ... Richstone, D. (1997). *AJ*, *114*, 1771.

Fabian, A. C. (2012). *Ann. Rev. A&A*, *50*, 455-489.

Fan, X., Narayanan, V. K., Lupton, R. H., Strauss, M. A., Knapp, G. R., Becker, R. H., ... York, D. G. (2001). *AJ*, *122*, 2833-2849.

Favata, M., Hughes, S. A., & Holz, D. E. (2004). *ApJL*, *607*, L5-L8.

Fialkov, A., & Barkana, R. (2014). *M.N.R.A.S.*, *445*, 213-224.

Fialkov, A., Barkana, R., & Visbal, E. (2014). *Nat.*, *506*, 197-199.

Fitchett, M. J., & Detweiler, S. (1984). *M.N.R.A.S.*, *211*, 933-942.

Fragos, T., Lehmer, B., Tremmel, M., Tzanavaris, P., Basu-Zych, A., Belczynski, K., ... Zezas, A. (2013). *ApJ*, *764*, 41.

Fragos, T., Lehmer, B. D., Naoz, S., Zezas, A., & Basu-Zych, A. (2013). *ApJL*, *776*, L31.

Frank, J., King, A., & Raine, D. J. (2002). *Accretion Power in Astrophysics*.

Fregeau, J. M., Cheung, P., Portegies Zwart, S. F., & Rasio, F. A. (2004). *M.N.R.A.S.*, *352*, 1-19.

Fryer, C. L., & Heger, A. (2005). *ApJ*, *623*, 302-313.

Furlanetto, S. R. (2006). *M.N.R.A.S.*, *371*, 867-878.

Gies, D. R., & Bolton, C. T. (1986). *ApJ Supp.*, *61*, 419-454.

Gilfanov, M., Grimm, H.-J., & Sunyaev, R. (2004). *M.N.R.A.S.*, *347*, L57-L60.

Gitti, M., Giroletti, M., Giovannini, G., Feretti, L., & Liuzzo, E. (2013). *A&A*, *557*, L14.

Gnedin, N. Y. (2000). *ApJ*, *542*, 535-541.

González, J. A., Hannam, M., Sperhake, U., Brügmann, B., & Husa, S. (2007). *Physical Review Letters*, *98*(23), 231101.

Graham, A. W., Erwin, P., Trujillo, I., & Asensio Ramos, A. (2003). *AJ*, *125*, 2951-2963.

Graham, A. W., & Guzmán, R. (2003). *AJ*, *125*, 2936-2950.

Greif, T. H., Bromm, V., Clark, P. C., Glover, S. C. O., Smith, R. J., Klessen, R. S., ... Springel, V. (2012). *M.N.R.A.S.*, *424*, 399-415.

Greif, T. H., Springel, V., White, S. D. M., Glover, S. C. O., Clark, P. C., Smith, R. J., ... Bromm, V. (2011). *ApJ*, *737*, 75.

Grimm, H.-J., Gilfanov, M., & Sunyaev, R. (2003). *M.N.R.A.S.*, *339*, 793-809.

Grindlay, J. E. (2009). In *American astronomical society meeting abstracts #213* (Vol. 41, p. #217.03).

Gualandris, A., & Merritt, D. (2008). *ApJ*, *678*, 780-797.

Gualandris, A., & Merritt, D. (2012a). *ApJ*, *744*, 74.

Gualandris, A., & Merritt, D. (2012b). *ApJ*, *744*, 74.

Gualandris, A., Portegies Zwart, S., & Eggleton, P. P. (2004). *M.N.R.A.S.*, *350*, 615-626.

Gualandris, A., Read, J. I., Dehnen, W., & Bortolas, E. (2017). *M.N.R.A.S.*, *464*, 2301-2310.

Guedes, J., Madau, P., Mayer, L., & Callegari, S. (2011). *ApJ*, *729*, 125.

Haiman, Z. (2011). *Nat.*, *472*, 47-48.

Haiman, Z. (2013). In T. Wiklind, B. Mobasher, & V. Bromm (Eds.), *Astrophysics and space science library* (Vol. 396, p. 293).

Haiman, Z., Kocsis, B., & Menou, K. (2009). *ApJ*, *700*, 1952-1969.

Haiman, Z., & Loeb, A. (2001). *ApJ*, *552*, 459-463.

Haiman, Z., Rees, M. J., & Loeb, A. (1996). *ApJ*, *467*, 522.

Heger, A., Fryer, C. L., Woosley, S. E., Langer, N., & Hartmann, D. H. (2003). *ApJ*, *591*, 288-300.

Heggie, D. C. (1975). *M.N.R.A.S.*, *173*, 729-787.

Heggie, D. C., & Rasio, F. A. (1996, October). The Effect of Encounters on the Eccentricity of Binaries in Clusters *M.N.R.A.S.*, *282*, 1064-1084. doi: 10.1093/mnras/282.3.1064

Herbig, G. H., & Terndrup, D. M. (1986). *ApJ*, *307*, 609-618.

Hills, J. G. (1975, October). Encounters between binary and single stars and their effect on the dynamical evolution of stellar systems *AJ*, *80*, 809-825. doi: 10.1086/111815

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., ... Wright, E. L. (2013). *ApJ Supp.*, *208*, 19.

Hirano, S., & Bromm, V. (2017). *M.N.R.A.S.*, *470*, 898-914.

Hirano, S., Hosokawa, T., Yoshida, N., Umeda, H., Omukai, K., Chiaki, G., & Yorke, H. W. (2014). *ApJ*, *781*, 60.

Hobbs, G., Archibald, A., Arzoumanian, Z., Backer, D., Bailes, M., Bhat, N. D. R., ... You, X. P. (2010). *Classical and Quantum Gravity*, *27*(8), 084013.

Hoffman, L., & Loeb, A. (2007). *M.N.R.A.S.*, *377*, 957-976.

Holley-Bockelmann, K., & Khan, F. M. (2015). *ApJ*, *810*, 139.

Hoogerwerf, R., de Bruijne, J. H. J., & de Zeeuw, P. T. (2000). *ApJL*, *544*, L133-L136.

Hopkins, P. F., & Quataert, E. (2010). *M.N.R.A.S.*, *407*, 1529-1564.

Hosokawa, T., Hirano, S., Kuiper, R., Yorke, H. W., Omukai, K., & Yoshida, N. (2015). *ArXiv e-prints*.

Hosokawa, T., Omukai, K., & Yorke, H. W. (2012). *ApJ*, *756*, 93.

Hosokawa, T., Omukai, K., Yoshida, N., & Yorke, H. W. (2011). *Science*, *334*, 1250.

Hughes, S. A. (2002). *M.N.R.A.S.*, *331*, 805-816.

Hummel, J. A., Stacy, A., Jeon, M., Oliveri, A., & Bromm, V. (2015). *M.N.R.A.S.*, *453*, 4136-4147.

Inayoshi, K., Haiman, Z., & Ostriker, J. P. (2016). *M.N.R.A.S.*, *459*, 3738-3755.

Inayoshi, K., & Tanaka, T. L. (2015). *M.N.R.A.S.*, *450*, 4350-4363.

Innanen, K. A., Tahtinen, L., & Valtonen, M. J. (1982). *AJ*, *87*, 1606.

Iwasawa, M., An, S., Matsubayashi, T., Funato, Y., & Makino, J. (2011). *ApJL*, *731*, L9.

Iwasawa, M., Funato, Y., & Makino, J. (2006). *ApJ*, *651*, 1059-1067.

Jaffe, A. H., & Backer, D. C. (2003). *ApJ*, *583*, 616-631.

Janka, H.-T. (2013). *M.N.R.A.S.*, *434*, 1355-1361.

Janssen, G., Hobbs, G., McLaughlin, M., Bassa, C., Deller, A., Kramer, M., ... Verbiest, J. P. W. (2015). *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 37.

Jenet, F. A., Hobbs, G. B., van Straten, W., Manchester, R. N., Bailes, M., Verbiest, J. P. W., ... Ord, S. M. (2006). *ApJ*, *653*, 1571-1576.

Jeon, M., Pawlik, A. H., Bromm, V., & Milosavljević, M. (2014). *M.N.R.A.S.*, *440*, 3778-3796.

Jeon, M., Pawlik, A. H., Greif, T. H., Glover, S. C. O., Bromm, V., Milosavljević, M., & Klessen, R. S. (2012). *ApJ*, *754*, 34.

Jiang, L., Cole, S., Sawala, T., & Frenk, C. S. (2015). *M.N.R.A.S.*, *448*, 1674-1686.

Just, A., & Peñarrubia, J. (2005). *A&A*, *431*, 861-877.

Kaaret, P. (2014). *M.N.R.A.S.*, *440*, L26-L30.

Kelley, L. Z., Blecha, L., & Hernquist, L. (2017). *M.N.R.A.S.*, *464*, 3131-3157.

Kesden, M., Sperhake, U., & Berti, E. (2010). *prd*, *81*(8), 084054.

Khan, F. M., Fiacconi, D., Mayer, L., Berczik, P., & Just, A. (2016). *ApJ*, *828*, 73.

Khan, F. M., Just, A., & Merritt, D. (2011). *ApJ*, *732*, 89.

Kharb, P., Lal, D. V., & Merritt, D. (2017). A candidate sub-parsec binary black hole in the Seyfert galaxy NGC 7674 *Nature Astronomy*, *1*, 727-733.

Khochfar, S., & Burkert, A. (2006). *A&A*, *445*, 403-412.

Kim, H., & Kim, W.-T. (2007). *ApJ*, *665*, 432-444.

Kim, W.-T. (2010). *ApJ*, *725*, 1069-1081.

Kocsis, B., & Sesana, A. (2011). *M.N.R.A.S.*, *411*, 1467-1479.

Kormendy, J., & Ho, L. C. (2013). *Ann. Rev. A&A*, *51*, 511-653.

Koushiappas, S. M., Bullock, J. S., & Dekel, A. (2004). *M.N.R.A.S.*, *354*, 292-304.

Kozai, Y. (1962). *AJ*, *67*, 591.

Kuhlen, M., & Madau, P. (2005). *M.N.R.A.S.*, *363*, 1069-1082.

Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N., & Cen, R. (2015). *ApJ*, *799*, 178.

Kulkarni, G., & Loeb, A. (2012). *M.N.R.A.S.*, *422*, 1306-1323.

Kupi, G., Amaro-Seoane, P., & Spurzem, R. (2006). *M.N.R.A.S.*, *371*, L45-L49.

Lajoie, C.-P., & Sills, A. (2011). *ApJ*, *726*, 67.

Latif, M. A., Bovino, S., Grassi, T., Schleicher, D. R. G., & Spaans, M. (2015). *M.N.R.A.S.*, *446*, 3163-3177.

Latif, M. A., Schleicher, D. R. G., & Schmidt, W. (2014). *M.N.R.A.S.*, *440*, 1551-1561.

Latif, M. A., Schleicher, D. R. G., Schmidt, W., & Niemeyer, J. (2013). *M.N.R.A.S.*, *433*, 1607-1618.

Lattimer, J. M., & Prakash, M. (2005). *Physical Review Letters*, *94*(11), 111101.

Lauer, T. R., Faber, S. M., Gebhardt, K., Richstone, D., Tremaine, S., Ajhar, E. A., ... Siopis, C. (2005). *AJ*, *129*, 2138-2185.

Leigh, N. W. C., Stone, N. C., Geller, A. M., Shara, M. M., Muddu, H., Solano-Oropeza, D., & Thomas, Y. (2016). *M.N.R.A.S.*, *463*, 3311-3325.

Lentati, L., Taylor, S. R., Mingarelli, C. M. F., Sesana, A., Sanidas, S. A., Vecchio, A., ... Verbiest, J. P. W. (2015). *M.N.R.A.S.*, *453*, 2576-2598.

Levine, R., Gnedin, N. Y., Hamilton, A. J. S., & Kravtsov, A. V. (2008). *ApJ*, *678*, 154-167.

Lewin, W. H. G., van Paradijs, J., & van den Heuvel, E. P. J. (1997). .

Lin, Y.-T., Stanford, S. A., Eisenhardt, P. R. M., Vikhlinin, A., Maughan, B. J., & Kravtsov, A. (2012). *ApJL*, *745*, L3.

Liu, Li, S., & Komossa, S. (2014). *ApJ*, *786*, 103.

Liu, Shen, Y., Bian, F., Loeb, A., & Tremaine, S. (2014). *ApJ*, *789*, 140.

Lodato, G., & Natarajan, P. (2006). *M.N.R.A.S.*, *371*, 1813-1823.

Lodato, G., Nayakshin, S., King, A. R., & Pringle, J. E. (2009). *M.N.R.A.S.*, *398*, 1392-1402.

Lousto, C. O., Campanelli, M., Zlochower, Y., & Nakano, H. (2010). *Classical and Quantum Gravity*, *27*(11), 114006.

Lousto, C. O., & Nakano, H. (2008). *Classical and Quantum Gravity*, *25*(19), 195019.

Lousto, C. O., & Zlochower, Y. (2011). *Physical Review Letters*, *107*(23), 231102.

Lousto, C. O., & Zlochower, Y. (2013). *prd*, *87*(8), 084027.

Lousto, C. O., Zlochower, Y., Dotti, M., & Volonteri, M. (2012). *prd*, *85*(8), 084015.

Lubow, S. H., & Ida, S. (2010). In S. Seager (Ed.), *Exoplanets* (p. 347-371).

Lupi, A., Haardt, F., Dotti, M., Fiacconi, D., Mayer, L., & Madau, P. (2016). *M.N.R.A.S.*, *456*, 2993-3003.

Ma, Q., Maio, U., Ciardi, B., & Salvaterra, R. (2015). *M.N.R.A.S.*, *449*, 3006-3014.

MacFadyen, A. I., & Woosley, S. E. (1999). *ApJ*, *524*, 262-289.

Machacek, M. E., Bryan, G. L., & Abel, T. (2003). *M.N.R.A.S.*, *338*, 273-286.

Machida, M. N., & Doi, K. (2013). *M.N.R.A.S.*, *435*, 3283-3305.

Madau, P., Haardt, F., & Dotti, M. (2014). *ApJL*, *784*, L38.

Madau, P., & Rees, M. J. (2001). *ApJL*, *551*, L27-L30.

Madau, P., Rees, M. J., Volonteri, M., Haardt, F., & Oh, S. P. (2004). *ApJ*, *604*, 484-494.

Magorrian, J., Tremaine, S., Gebhardt, K., Richstone, D., & Faber, S. (1996). In *American astronomical society meeting abstracts* (Vol. 28, p. 1423).

Maio, U., & Viel, M. (2015). *M.N.R.A.S.*, *446*, 2760-2775.

Manchester, R. N., Hobbs, G., Bailes, M., Coles, W. A., van Straten, W., Keith, M. J., ... You, X. P. (2013). *Publ. Astron. Soc. Australia*, *30*, e017.

Marigo, P., Chiosi, C., & Kudritzki, R.-P. (2003). *A&A*, *399*, 617-630.

Mayer, L. (2013). *Classical and Quantum Gravity*, *30*(24), 244008.

Mayer, L., Kazantzidis, S., Madau, P., Colpi, M., Quinn, T., & Wadsley, J. (2007). *Science*, *316*, 1874.

McGreer, I. D., Becker, R. H., Helfand, D. J., & White, R. L. (2006). *ApJ*, *652*, 157-162.

McWilliams, S. T., Ostriker, J. P., & Pretorius, F. (2014). *ApJ*, *789*, 156.

Menou, K., Haiman, Z., & Narayanan, V. K. (2001). *ApJ*, *558*, 535-542.

Merloni, A., & Heinz, S. (2008). *M.N.R.A.S.*, *388*, 1011-1030.

Merritt, D. (2006). *ApJ*, *648*, 976-986.

Merritt, D. (2013). *Dynamics and Evolution of Galactic Nuclei*.

Merritt, D., Alexander, T., Mikkola, S., & Will, C. M. (2010). *prd*, *81*(6), 062002.

Merritt, D., & Milosavljević, M. (2005). *Living Reviews in Relativity*, *8*, 8.

Merritt, D., Milosavljević, M., Favata, M., Hughes, S. A., & Holz, D. E. (2004). *ApJL*, *607*, L9-L12.

Merritt, D., & Valluri, M. (1996). *ApJ*, *471*, 82.

Merritt, D., & Wang, J. (2005). *ApJL*, *621*, L101-L104.

Mesinger, A., Perna, R., & Haiman, Z. (2005). *ApJ*, *623*, 1-10.

Migliari, S., & Fender, R. P. (2006). *M.N.R.A.S.*, *366*, 79-91.

Mikkola, S. (1983). *M.N.R.A.S.*, *203*, 1107-1121.

Milosavljević, M., Bromm, V., Couch, S. M., & Oh, S. P. (2009). *ApJ*, *698*, 766-780.

Milosavljević, M., & Merritt, D. (2001). *ApJ*, *563*, 34-62.

Milosavljević, M., & Merritt, D. (2003). In J. M. Centrella (Ed.), *The astrophysics of gravitational wave sources* (Vol. 686, p. 201-210).

Milosavljević, M., Merritt, D., Rest, A., & van den Bosch, F. C. (2002). *M.N.R.A.S.*, *331*, L51-L55.

Mineo, S., Gilfanov, M., & Sunyaev, R. (2012). *M.N.R.A.S.*, *419*, 2095-2115.

Mirabel, I. F., Dijkstra, M., Laurent, P., Loeb, A., & Pritchard, J. R. (2011). *A&A*, *528*, A149.

Mirza, M. A., Tahir, A., Khan, F. M., Holley-Bockelmann, H., Baig, A. M., Berczik, P., & Chishtie, F. (2017). *M.N.R.A.S.*, *470*, 940-947.

Mortlock, D. J., Warren, S. J., Venemans, B. P., Patel, M., Hewett, P. C., McMahon, R. G., ... Röttgering, H. J. A. (2011). *Nat.*, *474*, 616-619.

Moster, B. P., Naab, T., & White, S. D. M. (2013). *M.N.R.A.S.*, *428*, 3121-3138.

Naoz, S., & Barkana, R. (2007). *M.N.R.A.S.*, *377*, 667-676.

Narayan, R., Paczynski, B., & Piran, T. (1992). *ApJL*, *395*, L83-L86.

Navarro, J. F., Frenk, C. S., & White, S. D. M. (1997). *ApJ*, *490*, 493-508.

Oh, S. P. (2001). *ApJ*, *553*, 499-512.

Oh, S. P., & Haiman, Z. (2002). *ApJ*, *569*, 558-572.

Ostriker, E. C. (1999a). *ApJ*, *513*, 252-258.

Ostriker, E. C. (1999b). *ApJ*, *513*, 252-258.

Pacucci, F., & Ferrara, A. (2015). *M.N.R.A.S.*, *448*, 104-118.

Pacucci, F., Ferrara, A., Volonteri, M., & Dubus, G. (2015). *ArXiv e-prints*.

Pacucci, F., Volonteri, M., & Ferrara, A. (2015). *M.N.R.A.S.*, *452*, 1922-1933.

Pâris, I., Petitjean, P., Aubourg, E., Myers, A. D., Streblyanska, A., Lyke, B. W., ... Zhao, G.-B. (2017). *ArXiv e-prints*.

Park, K., & Ricotti, M. (2012). *ApJ*, *747*, 9.

Perets, H. B., & Alexander, T. (2008). *ApJ*, *677*, 146-159.

Perets, H. B., Hopman, C., & Alexander, T. (2007). *ApJ*, *656*, 709-720.

Perna, R., Duffell, P., Cantiello, M., & MacFadyen, A. I. (2014). *ApJ*, *781*, 119.

Persic, M., Rephaeli, Y., Braito, V., Cappi, M., Della Ceca, R., Franceschini, A., & Gruber, D. E. (2004). *A&A*, *419*, 849-862.

Peters, P. C. (1964). *Physical Review*, *136*, 1224-1232.

Peters, P. C., & Mathews, J. (1963). *Physical Review*, *131*, 435-440.

Pezzulli, E., Valiante, R., & Schneider, R. (2016). *M.N.R.A.S.*, *458*, 3047-3059.

Pfalzner, S. (2009). *A&A*, *498*, L37-L40.

Phinney, E. S. (2001). *ArXiv Astrophysics e-prints*.

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., ... et al. (2015). *ArXiv e-prints*.

Poon, M. Y., & Merritt, D. (2002). *ApJL*, *568*, L89-L92.

Poon, M. Y., & Merritt, D. (2004). *ApJ*, *606*, 774-787.

Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. (2010). *Ann. Rev. A&A*, *48*, 431-493.

Poveda, A., Ruiz, J., & Allen, C. (1967). *Boletin de los Observatorios Tonantzintla y Tacubaya*, *4*, 86-90.

Power, C., Wynn, G. A., Combet, C., & Wilkinson, M. I. (2009). *M.N.R.A.S.*, *395*, 1146-1152.

Preto, M., Berentzen, I., Berczik, P., & Spurzem, R. (2011). *ApJL*, *732*, L26.

Prieto, J., & Escala, A. (2016). *M.N.R.A.S.*.

Pritchard, J. R., & Furlanetto, S. R. (2007). *M.N.R.A.S.*, *376*, 1680-1694.

Pritchard, J. R., & Loeb, A. (2008). *prd*, *78*(10), 103511.

Quinlan, G. D. (1996). *New Astron.*, *1*, 35-56.

Rasskazov, A., & Merritt, D. (2017). *ApJ*, *837*, 135.

Ravi, V., Wyithe, J. S. B., Hobbs, G., Shannon, R. M., Manchester, R. N., Yardley, D. R. B., & Keith, M. J. (2012). *ApJ*, *761*, 84.

Ravi, V., Wyithe, J. S. B., Shannon, R. M., Hobbs, G., & Manchester, R. N. (2014). *M.N.R.A.S.*, *442*, 56-68.

Rees, M. J., & Ostriker, J. P. (1977). *M.N.R.A.S.*, *179*, 541-559.

Regan, J. A., & Haehnelt, M. G. (2009). *M.N.R.A.S.*, *396*, 343-353.

Regan, J. A., Johansson, P. H., & Haehnelt, M. G. (2014). *M.N.R.A.S.*, *439*, 1160-1175.

Regan, J. A., Johansson, P. H., & Wise, J. H. (2016). *M.N.R.A.S., submitted, e-print arXiv:1603.06612*.

Regös, E., Bailey, V. C., & Mardling, R. (2005). *M.N.R.A.S.*, *358*, 544-550.

Repetto, S., Davies, M. B., & Sigurdsson, S. (2012). *M.N.R.A.S.*, *425*, 2799-2809.

Ricotti, M., & Ostriker, J. P. (2004a). *M.N.R.A.S.*, *352*, 547-562.

Ricotti, M., & Ostriker, J. P. (2004b). *M.N.R.A.S.*, *352*, 547-562.

Ripamonti, E., Mapelli, M., & Zaroubi, S. (2008). *M.N.R.A.S.*, *387*, 158-172.

Rodriguez, C., Taylor, G. B., Zavala, R. T., Peck, A. B., Pollack, L. K., & Romani, R. W. (2006). *ApJ*, *646*, 49-60.

Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., Sijacki, D., Pillepich, A., Sales, L. V., ... Hernquist, L. (2015). *M.N.R.A.S.*, *449*, 49-64.

Roedig, C., Dotti, M., Sesana, A., Cuadra, J., & Colpi, M. (2011). *M.N.R.A.S.*, *415*, 3033-3041.

Ryu, T., Leigh, N. W. C., & Perna, R. (2017a). *M.N.R.A.S.*, *467*, 4447-4461.

Ryu, T., Leigh, N. W. C., & Perna, R. (2017b). *M.N.R.A.S.*, *470*, 2-19.

Ryu, T., Leigh, N. W. C., & Perna, R. (2017c). *M.N.R.A.S.*, *470*, 3049-3067.

Ryu, T., Leigh, N. W. C., & Perna, R. (2017d, September). An analytic method for identifying dynamically formed runaway stars *M.N.R.A.S.*, *470*, 2-19. doi: 10.1093/mnras/stx1156

Ryu, T., Perna, R., Haiman, Z., Ostriker, J. P., & Stone, N. C. (2018). *M.N.R.A.S.*, *473*, 3410-3433.

Ryu, T., Tanaka, T. L., & Perna, R. (2016). *M.N.R.A.S.*, *456*, 223-238.

Ryu, T., Tanaka, T. L., Perna, R., & Haiman, Z. (2016). *M.N.R.A.S.*, *460*, 4122-4134.

Salvaterra, R., Campana, S., Chincarini, G., Covino, S., & Tagliaferri, G. (2008). *M.N.R.A.S.*, *385*, 189-194.

Salvaterra, R., Haardt, F., Volonteri, M., & Moretti, A. (2012). *A&A*, *545*, L6.

Sánchez-Salcedo, F. J., & Brandenburg, A. (2001). *M.N.R.A.S.*, *322*, 67-78.

Schaerer, D. (2002). In M. Chávez, A. Bressan, A. Buzzoni, & D. Mayya (Eds.), *New quests in stellar astrophysics: the link between stars and cosmology* (Vol. 274, p. 185-188).

Schanne, S., Paul, J., Wei, J., Zhang, S. ., Basa, S., Atteia, J. ., ... Osborne, J. (2010). *ArXiv e-prints*.

Schleicher, D. R. G., Spaans, M., & Glover, S. C. O. (2010). *ApJL*, *712*, L69-L72.

Schnittman, J. D. (2007). *ApJL*, *667*, L133-L136.

Seager, S. (2010). *Exoplanets*.

Sepinsky, J. F., Willems, B., & Kalogera, V. (2007). *ApJ*, *660*, 1624-1635.

Sepinsky, J. F., Willems, B., Kalogera, V., & Rasio, F. A. (2009). *ApJ*, *702*, 1387-1392.

Sepinsky, J. F., Willems, B., Kalogera, V., & Rasio, F. A. (2010). *ApJ*, *724*, 546-558.

Sesana, A. (2010). *ApJ*, *719*, 851-864.

Sesana, A. (2013). *M.N.R.A.S.*, *433*, L1-L5.

Sesana, A. (2015). In C. F. Sopuerta (Ed.), *Gravitational wave astrophysics* (Vol. 40, p. 147).

Sesana, A., Gualandris, A., & Dotti, M. (2011). *M.N.R.A.S.*, *415*, L35-L39.

Sesana, A., Shankar, F., Bernardi, M., & Sheth, R. K. (2016). *M.N.R.A.S.*, *463*, L6-L11.

Sesana, A., Vecchio, A., & Colacino, C. N. (2008). *M.N.R.A.S.*, *390*, 192-209.

Sesana, A., Vecchio, A., & Volonteri, M. (2009). *M.N.R.A.S.*, *394*, 2255-2265.

Sesana, A., Volonteri, M., & Haardt, F. (2007). *M.N.R.A.S.*, *377*, 1711-1716.

Shakura, N. I., & Sunyaev, R. A. (1973). *A&A*, *24*, 337-355.

Shang, C., Bryan, G. L., & Haiman, Z. (2010). *M.N.R.A.S.*, *402*, 1249-1262.

Shankar, F., Bernardi, M., Sheth, R. K., Ferrarese, L., Graham, A. W., Savorgnan, G., ... Lapi, A. (2016). *M.N.R.A.S.*.

Shankar, F., Crocce, M., Miralda-Escudé, J., Fosalba, P., & Weinberg, D. H. (2010). *ApJ*, *718*, 231-250.

Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. (2009). *ApJ*, *690*, 20-41.

Shannon, R. M., Ravi, V., Coles, W. A., Hobbs, G., Keith, M. J., Manchester, R. N., ... Wang, J.-B. (2013). *Science*, *342*, 334-337.

Shannon, R. M., Ravi, V., Lentati, L. T., Lasky, P. D., Hobbs, G., Kerr, M., ... Zhu, X.-J. (2015). *Science*, *349*, 1522-1525.

Shapiro, S. L. (2005). *ApJ*, *620*, 59-68.

Shen, Y., Strauss, M. A., Oguri, M., Hennawi, J. F., Fan, X., Richards, G. T., ... Knapp, G. R. (2007). *AJ*, *133*, 2222-2241.

Shields, G. A., Rosario, D. J., Smith, K. L., Bonning, E. W., Salviander, S., Kalirai, J. S., ... Marshall, P. J. (2009). *ApJ*, *707*, 936-941.

Shlosman, I., Frank, J., & Begelman, M. C. (1989). *Nat.*, *338*, 45-47.

Shtykovskiy, P. E., & Gilfanov, M. R. (2007). *Astronomy Letters*, *33*, 437-454.

Shu, F. H. (1977). *ApJ*, *214*, 488-497.

Sijacki, D., Springel, V., Di Matteo, T., & Hernquist, L. (2007). *M.N.R.A.S.*, *380*, 877-900.

Silk, J. (1977). *ApJ*, *211*, 638-648.

Silk, J., & Rees, M. J. (1998). *A&A*, *331*, L1-L4.

Sipior, M. S., Eracleous, M., & Sigurdsson, S. (2003). *ArXiv Astrophysics e-prints*.

Spaans, M., & Silk, J. (2006). *ApJ*, *652*, 902-906.

Spinnato, P. F., Fellhauer, M., & Portegies Zwart, S. F. (2003). *M.N.R.A.S.*, *344*, 22-32.

Springel, V., Di Matteo, T., & Hernquist, L. (2005). *M.N.R.A.S.*, *361*, 776-794.

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., ... Pearce, F. (2005). *Nat.*, *435*, 629-636.

Spruit, H. C. (2002). *A&A*, *381*, 923-932.

Stacy, A., & Bromm, V. (2013). *M.N.R.A.S.*, *433*, 1094-1107.

Stacy, A., Bromm, V., & Lee, A. T. (2016). *M.N.R.A.S.*, *462*, 1307-1328.

Stacy, A., Greif, T. H., & Bromm, V. (2010). *M.N.R.A.S.*, *403*, 45-60.

Stacy, A., Greif, T. H., & Bromm, V. (2012). *M.N.R.A.S.*, *422*, 290-309.

Stinson, G., Seth, A., Katz, N., Wadsley, J., Governato, F., & Quinn, T. (2006). *M.N.R.A.S.*, *373*, 1074-1090.

Stone, & Ostriker, J. P. (2015). *ApJL*, *806*, L28.

Stone, R. C. (1979). *ApJ*, *232*, 520-530.

Strateva, I. V., Strauss, M. A., Hao, L., Schlegel, D. J., Hall, P. B., Gunn, J. E., ... Nichol, R. C. (2003). *AJ*, *126*, 1720-1749.

Sugimura, K., Omukai, K., & Inoue, A. K. (2014). *M.N.R.A.S.*, *445*, 544-553.

Susa, H., Hasegawa, K., & Tominaga, N. (2014). *ApJ*, *792*, 32.

Sweatman, W. L. (2007). *M.N.R.A.S.*, *377*, 459-464.

Syer, D., & Clarke, C. J. (1995). *M.N.R.A.S.*, *277*, 758-766.

Taffoni, G., Mayer, L., Colpi, M., & Governato, F. (2003). *M.N.R.A.S.*, *341*, 434-448.

Tagawa, H., Umemura, M., & Gouda, N. (2016). *MNRAS, submitted; e-print arXiv:1602.08767*.

Tagawa, H., Umemura, M., Gouda, N., Yano, T., & Yamai, Y. (2015). *M.N.R.A.S.*, *451*, 2174-2184.

Tan, J. C., & Blackman, E. G. (2004). *ApJ*, *603*, 401-413.

Tan, J. C., & McKee, C. F. (2004). *ApJ*, *603*, 383-400.

Tanaka. (2014). *Classical and Quantum Gravity*, *31*(24), 244005.

Tanaka, & Haiman, Z. (2009). *ApJ*, *696*, 1798-1822.

Tanaka, & Li, M. (2014). *M.N.R.A.S.*, *439*, 1092-1100.

Tanaka, O'Leary, R. M., & Perna, R. (2016). *M.N.R.A.S.*, *455*, 2619-2626.

Tanaka, Perna, R., & Haiman, Z. (2012a). *M.N.R.A.S.*, *425*, 2974-2987.

Tanaka, Perna, R., & Haiman, Z. (2012b). *M.N.R.A.S.*, *425*, 2974-2987.

Tang, Y., MacFadyen, A., & Haiman, Z. (2017). *ArXiv e-prints*.

Taylor, S. R., Vallisneri, M., Ellis, J. A., Mingarelli, C. M. F., Lazio, T. J. W., & van Haasteren, R. (2016). *ApJL*, *819*, L6.

Tegmark, M., Silk, J., Rees, M. J., Blanchard, A., Abel, T., & Palla, F. (1997). *ApJ*, *474*, 1.

The NANOGrav Collaboration, Arzoumanian, Z., Brazier, A., Burke-Spolaor, S., Chamberlin, S., Chatterjee, S., ... Zhu, W. (2015). *ApJ*, *813*, 65.

Thomas, J., Ma, C.-P., McConnell, N. J., Greene, J. E., Blakeslee, J. P., & Janish, R. (2016). *Nat.*, *532*, 340-342.

Thomas, J., Saglia, R. P., Bender, R., Erwin, P., & Fabricius, M. (2014). *ApJ*, *782*, 39.

Toma, K., Sakamoto, T., & Mészáros, P. (2011). *ApJ*, *731*, 127.

Treister, E., Schawinski, K., Urry, C. M., & Simmons, B. D. (2012). *ApJL*, *758*, L39.

Tremaine, S., Gebhardt, K., Bender, R., Bower, G., Dressler, A., Faber, S. M., ... Richstone, D. (2002). *ApJ*, *574*, 740-753.

Tremblay, S. E., Taylor, G. B., Ortiz, A. A., Tremblay, C. D., Helmboldt, J. F., & Romani, R. W. (2016). *M.N.R.A.S.*, *459*, 820-840.

Tsalmantza, P., Decarli, R., Dotti, M., & Hogg, D. W. (2011). *ApJ*, *738*, 20.

Turk, M. J., Abel, T., & O'Shea, B. (2009). *Science*, *325*, 601-.

Turon, C., et al. (Eds.). (1992). *ESA Special Publication* (Vol. 1136).

Valtonen, M., & Karttunen, H. (2006). *The Three-Body Problem*.

Valtonen, M., & Mikkola, S. (1991). *Ann. Rev. A&A*, *29*, 9-29.

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., ... van Zwieten, J. (2013). *A&A*, *556*, A2.

Van Wassenhove, S., Capelo, P. R., Volonteri, M., Dotti, M., Bellovary, J. M., Mayer, L., & Governato, F. (2014). *M.N.R.A.S.*, *439*, 474-487.

Vasiliev, E., Antonini, F., & Merritt, D. (2014). *ApJ*, *785*, 163.

Vasiliev, E., Antonini, F., & Merritt, D. (2015). *ApJ*, *810*, 49.

Venemans, B. P., Findlay, J. R., Sutherland, W. J., De Rosa, G., McMahon, R. G., Simcoe, R., ... Lewis, J. R. (2013). *ApJ*, *779*, 24.

Venkatesan, A., Giroux, M. L., & Shull, J. M. (2001). *ApJ*, *563*, 1-8.

Venkatesan, A., Tumlinson, J., & Shull, J. M. (2003). *ApJ*, *584*, 621-632.

Verbiest, J. P. W., Lentati, L., Hobbs, G., van Haasteren, R., Demorest, P. B., Janssen, G. H., ... Zhu, X.-J. (2016). *M.N.R.A.S.*, *458*, 1267-1288.

Visbal, E., Haiman, Z., & Bryan, G. L. (2014). *M.N.R.A.S.*, *445*, 1056-1063.

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., ... Hernquist, L. (2014). *Nat.*, *509*, 177-182.

Volonteri, M. (2010). *A&A Rev.*, *18*, 279-315.

Volonteri, M., Haardt, F., & Madau, P. (2003). *ApJ*, *582*, 559-573.

Volonteri, M., & Rees, M. J. (2005). *ApJ*, *633*, 624-629.

Volonteri, M., & Rees, M. J. (2006). *ApJ*, *650*, 669-678.

Volonteri, M., Silk, J., & Dubus, G. (2015). *ApJ*, *804*, 148.

Voytek, T. C., Natarajan, A., Jáuregui García, J. M., Peterson, J. B., & López-Cruz, O. (2014). *ApJL*, *782*, L9.

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. (2002). *ApJ*, *568*, 52-70.

Wetzel, A. R. (2011). *M.N.R.A.S.*, *412*, 49-58.

Willott, C. J., Delorme, P., Omont, A., Bergeron, J., Delfosse, X., Forveille, T., ... Cox, P. (2007). *AJ*, *134*, 2435-2450.

Willott, C. J., Delorme, P., Reylé, C., Albert, L., Bergeron, J., Crampton, D., ... Schade, D. (2009). *AJ*, *137*, 3541-3547.

Willott, C. J., Delorme, P., Reylé, C., Albert, L., Bergeron, J., Crampton, D., ... Schade, D. (2010). *AJ*, *139*, 906-918.

Willott, C. J., McLure, R. J., & Jarvis, M. J. (2003). *ApJL*, *587*, L15-L18.

Wise, J. H., & Abel, T. (2007). *ApJ*, *665*, 899-910.

Wise, J. H., & Abel, T. (2008). *ApJ*, *685*, 40-56.

Wise, J. H., Turk, M. J., & Abel, T. (2008). *ApJ*, *682*, 745-757.

Wyithe, J. S. B., & Loeb, A. (2003). *ApJ*, *590*, 691-706.

Wyithe, J. S. B., & Loeb, A. (2012). *M.N.R.A.S.*, *425*, 2892-2902.

Xu, H., Wise, J. H., & Norman, M. L. (2013). *ApJ*, *773*, 83.

Yoo, J., & Miralda-Escudé, J. (2004). *ApJL*, *614*, L25-L28.

Yoon, S.-C., Langer, N., & Norman, C. (2006). *A&A*, *460*, 199-208.

Yoshida, N., Omukai, K., Hernquist, L., & Abel, T. (2006). *ApJ*, *652*, 6-25.

Yu, Q. (2002). *M.N.R.A.S.*, *331*, 935-958.

Zhang, W., Woosley, S. E., & Heger, A. (2008). *ApJ*, *679*, 639-654.

A Mergers and gravitational wave (GW) recoil kicks

When two BHs merge under user-specified criteria (i.e. GW decay time < orbital period), the remnant BH gets a kick due to anisotropic emission of gravitational waves (Bekenstein, 1973; Fitchett & Detweiler, 1984; Favata et al., 2004).

We implement the effects of the recoil kick for BHs and take into account the mass loss to gravitational radiation for the remnant BH using the analytic formulae with the best-fit values given in Lousto et al. (2010). For the recoil kick, the fitting formula can be written as follows,

$$\boldsymbol{v}_{\rm recoil}(q,\boldsymbol{\alpha}) = v_{\rm m}\hat{e}_1 + v_\perp(\cos\xi\ \hat{e}_1 + \sin\xi\ \hat{e}_2) + v_\parallel\hat{n}_\parallel, \tag{106}$$

$$\begin{aligned} v_{\rm m} &= A\frac{\eta^2(1-q)}{1+q}[1+B\eta] \\ v_\perp &= H\frac{\eta^2}{1+q}(1+B_{\rm H}\eta)(\alpha_2^\parallel - q\alpha_1^\parallel) \\ v_\parallel &= K\frac{\eta^2}{1+q}(1+B_{\rm K}\eta)(\alpha_2^\perp - q\alpha_1^\perp)\cos(\Theta_\Delta - \Theta_0)\,, \end{aligned} \tag{107}$$

where q is the mass ratio of two BHs in binaries, $M_1/M_2(<1)$, $\eta = q/(1+q)^2$ and $\boldsymbol{\alpha}_i = \boldsymbol{S}_i/M_i^2$ is the intrinsic spin of BH i and the indices $\perp$ and $\parallel$ refer to perpendicular and parallel to the orbital angular momentum, respectively. $\hat{e}_1$ and $\hat{e}_2$ are orthogonal unit vectors in the orbital plane, ξ measures the angle between the unequal mass and spin contribution to the recoil velocity in the orbital plane. $\Theta_\Delta - \Theta_0$ is the angle difference between the in-plane component and the infall direction at merger. Adopting their findings, we take $A = 1.2\times10^4$ km s^{-1}, $B = -0.93$, $H = 6.9\times10^3$ km s^{-1}, $B_{H,K} = 0$, $K = 6.0\times10^4$ km s^{-1} and $\xi = 145^\circ$.

Using the same parameters drawn for the recoil velocities, we estimate remnant masses using Eq. (4) up to the leading order and Eq. (5) in Lousto et al. (2010). For two BHs of M_1 and M_2 of a binary, the remnant mass $M_{\rm remnant}$ is expressed as follows,

$$\frac{\Delta M}{M_1+M_2} = \eta\tilde{E}_{ISCO}, \tag{108}$$

$$\begin{aligned} \tilde{E}_{ISCO} &= (1-\frac{\sqrt{8}}{3}) + 0.103803\eta \\ &+ \frac{1}{36\sqrt{3}(1+q)^2}\left[q(1+2q)\alpha_1^\parallel + (2+q)\alpha_2^\perp\right] \\ &- \frac{5}{324\sqrt{2}(1+q)^2}\Big[\boldsymbol{\alpha}_2^2 - 3(\alpha_2^\parallel)^2 - 2q(\boldsymbol{\alpha}_1\cdot\boldsymbol{\alpha}_2 - 3\alpha_1^\parallel\alpha_2^\parallel) \\ &+ q^2(\boldsymbol{\alpha}_1^2 - 3(\alpha_1^\parallel)^2)\Big] \end{aligned} \tag{109}$$

$$M_{\rm remnant} = M_1 + M_2 - \Delta M \tag{110}$$

total stellar mass $M_\star$	$\left(\frac{M_\star}{10^{11}\ \mathrm{M}_\odot}\right) = 2.00\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)$
	$\left(\frac{M_\star}{10^{11}\ \mathrm{M}_\odot}\right) = 1.85\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{0.86}$
stellar dispersion σ	$\left(\frac{\sigma}{200\ \mathrm{km}s^{-1}}\right) = 1.34\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)^{0.26}$
	$\left(\frac{\sigma}{200\ \mathrm{km}s^{-1}}\right) = 1.31\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{0.23}$
half-mass radius $r_{\rm h}$	$\left(\frac{r_{\rm h}}{\mathrm{kpc}}\right) = 5.34\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)^{0.47}$
	$\left(\frac{r_{\rm h}}{\mathrm{kpc}}\right) = 5.14\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{0.41}$
core radius $r_{\rm c}$	$\left(\frac{r_{\rm c}}{\mathrm{pc}}\right) = 0.92\left(\frac{M_{\rm DM}}{10^{11}\ \mathrm{M}_\odot}\right)^{0.99}$ or $\left(\frac{r_{\rm c}}{\mathrm{kpc}}\right) = 0.089\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)^{0.99}$
	$\left(\frac{r_{\rm c}}{\mathrm{pc}}\right) = 1.56\left(\frac{M_{\rm BH}}{10^{7}\ \mathrm{M}_\odot}\right)^{0.86}$ or $\left(\frac{r_{\rm c}}{\mathrm{kpc}}\right) = 0.082\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{0.86}$
core density $\rho_{\rm c}$	$\left(\frac{\rho_{\rm c}}{\mathrm{M}_\odot\ \mathrm{pc}^{-3}}\right) = 240\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)^{-1.46}$
	$\left(\frac{\rho_{\rm c}}{\mathrm{M}_\odot\ \mathrm{pc}^{-3}}\right) = 270\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{-1.26}$
core mass $M_{\rm c}$	$\left(\frac{M_{\rm c}}{10^{9}\ \mathrm{M}_\odot}\right) = 0.453\left(\frac{M_{\rm DM}}{10^{13}\ \mathrm{M}_\odot}\right)^{1.52} = 0.158\left(\frac{M_\star}{10^{11}\ \mathrm{M}_\odot}\right)^{1.52}$
	$\left(\frac{M_{\rm c}}{10^{9}\ \mathrm{M}_\odot}\right) = 0.403\left(\frac{M_{\rm BH}}{10^{9}\ \mathrm{M}_\odot}\right)^{1.31}$

Table 13: The variables relevant for the model galaxy in this work in terms of BH mass $M_{\rm BH}$ and DM halo mass $M_{\rm DM}$, derived using the scaling relations 1 to 4.

B Scaling relations

We provide in Table 13 the scaling relations between the relevant variables in our model in terms of $M_{\rm BH}$ (as well as $M_{\rm DM}$), derived with the four scaling relations 1-4. We show derivations for some of the exponents in the table which are less immediate.

1. *Relations of σ*:

   Using the scaling relation 3 and the relation between $M_{\rm BH}$ and $M_{\rm DM}$ (same as the scaling relation 2),

   $$\sigma \sim M_{\rm BH}^{1/4.38} \sim M_{\rm DM}^{1.16/4.38} = M_{\rm DM}^{0.26}. \tag{111}$$

2. *Relations of $r_{\rm h}$*:

   Combining the relation $r_{\rm h} \sim M_\star/\sigma^2$ and the scaling relation 1 and 3 (or Equation 111 derived above),

   $$r_{\rm h} \sim M_\star \sigma^{-2} \sim M_{\rm DM}\left(M_{\rm DM}^{-0.26}\right)^2 \sim M_{\rm DM}^{0.47} \sim M_\star^{0.47}. \tag{112}$$

And the relation $M_{\rm DM} \sim M_{\rm BH}^{0.86}$ gives,

$$r_{\rm h} \sim M_{\rm DM}^{0.47} \sim M_{\rm BH}^{0.41}. \tag{113}$$

3. *Relations of* $r_{\rm c}$:

From the scaling relation 4 we find

$$r_{\rm c} \sim M_{\rm BH}^{0.86} \sim M_{\rm DM}^{0.86\times1.16} \sim M_{\rm DM}^{0.99} \sim M_{\star}^{0.99}. \tag{114}$$

4. *Relations of* $\rho_{\rm c}$:

Given Equation (5) in Stone & Ostriker (2015) and Equations 113 and 114, for $r_{\rm h} \gg r_{\rm c}$,

$$\rho_{\rm c} \sim M_{\star} r_{\rm h}^{-1} \left(r_{\rm c}\right)^{-2} \sim M_{\rm DM} M_{\rm DM}^{-0.41} \left(M_{\rm DM}^{-0.99}\right)^2 \sim M_{\rm DM}^{-1.46} \sim M_{\star}^{-1.46}. \tag{115}$$

The scaling relation 2 gives

$$\rho_{\rm c} \sim M_{\rm DM}^{-1.46} \sim M_{\rm BH}^{-1.26}. \tag{116}$$

5. *Relations of* $M_{\rm c}$:

Starting with Equation (6) in Stone & Ostriker (2015),

$$M_{\rm c} \sim r_{\rm c} \left(r_{\rm h}\right)^{-1} M_{\star}, \tag{117}$$

and inserting Equation 112 and 114 in to Equation 117 above, we arrive at the expressions

$$M_{\rm c} \sim r_{\rm c} \left(r_{\rm h}\right)^{-1} M_{\star} \sim M_{\rm DM}^{0.99} M_{\rm DM}^{-0.47} M_{\rm DM} \sim M_{\rm DM}^{1.52} \sim M_{\star}^{1.52}, \tag{118}$$

and

$$M_{\rm c} \sim M_{\rm BH}^{1.31}. \tag{119}$$

6. $r_{\rm c} - r_{\rm h}$ *relation*:

$$\frac{r_{\rm h}}{r_{\rm c}} = 63 \left(\frac{M_{\rm BH}}{10^9\ {\rm M}_\odot}\right)^{-0.45} = 60 \left(\frac{M_{\rm DM}}{10^{13}\ {\rm M}_\odot}\right)^{-0.52}, \tag{120}$$

or

$$\left(\frac{r_{\rm h}}{\rm kpc}\right) = 16.9 \left(\frac{r_{\rm c}}{\rm kpc}\right)^{0.48} \rightarrow \left(\frac{r_{\rm h}}{\rm kpc}\right) = 0.61 \left(\frac{r_{\rm c}}{\rm pc}\right)^{0.48}, \tag{121}$$

$$\left(\frac{r_{\rm c}}{\rm kpc}\right) = 0.0026 \left(\frac{r_{\rm h}}{\rm kpc}\right)^{2.10} \rightarrow \left(\frac{r_{\rm c}}{\rm pc}\right) = 2.6 \left(\frac{r_{\rm h}}{\rm kpc}\right)^{2.10}. \tag{122}$$