Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 3488–3503 (2022) https://doi.org/10.1093/mnras/stab3239
Advance Access publication 2021 November 10

Massive black hole evolution models confronting the n-Hz amplitude of the stochastic gravitational wave background

David Izquierdo-Villalba,[1,2]★ Alberto Sesana,[1] Silvia Bonoli[3,4] and Monica Colpi[1,2]

[1]*Dipartimento di Fisica 'G. Occhialini', Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[2]*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*
[3]*Donostia International Physics Centre (DIPC), Paseo Manuel de Lardizabal 4, E-20018 Donostia-San Sebastian, Spain*
[4]*IKERBASQUE, Basque Foundation for Science, E-48013 Bilbao, Spain*

Accepted 2021 November 4. Received 2021 November 4; in original form 2021 August 26

ABSTRACT

We estimate the amplitude of the nano-Hz stochastic gravitational wave background (GWB) resulting from an unresolved population of inspiralling massive black hole binaries (MBHBs). To this aim, we use the `L-Galaxies` semi-analytical model applied on top of the `Millennium` merger trees. The dynamical evolution of MBHBs includes dynamical friction, stellar and gas binary hardening, and gravitational wave (GW) feedback. At the frequencies proved by the *Pulsar Timing Array* experiments, our model predicts an amplitude of $\sim 1.2 \times 10^{-15}$ at $\sim 3 \times 10^{-8}$ Hz in agreement with current estimations. The contribution to the background comes primarily from equal-mass binaries with chirp masses above $10^8\,M_\odot$. We then consider the recently detected common red noise in NANOGrav, PPTA, and EPTA data, working under the hypothesis that it is indeed a stochastic GWB coming from MBHBs. By boosting the massive black hole growth via gas accretion, we show that our model can produce a signal with an amplitude $A \approx (2\text{–}3) \times 10^{-15}$. There are, however, difficulties in predicting this background level without mismatching key observational constraints such as the quasar bolometric luminosity functions or the local black hole mass function. This highlights how current and forthcoming GW observations can, for the first time, confront galaxy and black hole evolution models.

Key words: black hole physics – gravitational waves.

1 INTRODUCTION

Due to recent major advances in the observational studies of active galactic nuclei (AGN), evidence is growing that massive black holes (MBHs) heavier than $10^5\,M_\odot$ form in nature and power AGN activity at the centres of galaxies through gas accretion (Schmidt 1963; Hopkins, Richards & Hernquist 2007; Merloni & Heinz 2008; Ueda et al. 2014; Aird et al. 2015). The demographic study of AGN and the dynamics of stars and gas around the centre of nearby galaxies, further provided evidence that most (if not all) massive galaxies in the Universe host MBHs in their nuclei (Genzel & Townes 1987; Dressler & Richstone 1988; Kormendy 1988; Kormendy & Richstone 1992; Genzel, Hollenbach & Townes 1994; Salucci et al. 1999; Peterson et al. 2004; Vestergaard & Peterson 2006). Even more, the existing correlations between the mass of MBHs and key properties of their host galaxies hint for their co-evolution (Haehnelt & Rees 1993; Faber 1999; O'Dowd, Urry & Scarpa 2002; Häring & Rix 2004; Kormendy & Ho 2013; Savorgnan et al. 2016). Even though these findings sharpened our knowledge on the role of MBHs in the formation and evolution of galaxies, there is a need to contextualize galaxies and MBHs within the broad cosmological context. It is commonly accepted that the Universe behaves in a hierarchical way. Cosmic structure formed through the hierarchical assembly of dark matter (DM) haloes, and the galaxies observed nowadays assembled through mergers with smaller companions and accretion of matter from the cosmic filaments (White & Rees 1978; White & Frenk 1991; Haehnelt & Rees 1993; Kauffmann et al. 1999; Guo et al. 2011; Vogelsberger et al. 2014a,b; Schaye et al. 2015; Nelson et al. 2018; Pillepich et al. 2018). Consequently, the existence of MBHs at the centre of galaxies and the main role of mergers in the Universe, hint for the existence of *massive black hole binary systems* (MBHBs) that might have formed and coalesced throughout the whole Universe lifetime.

Discovering the population of MBHBs is compelling but detecting dual or binary AGN over a wide mass spectrum and redshift space is still a challenge (see, for a review De Rosa et al. 2019). An alternative avenue to discover MBHBs is provided by General Relativity. According to the theory, in fact, MBHBs are sources of gravitational waves (GWs), with frequencies ranging from above 10^{-9} Hz up to a few 10^{-2} Hz (Sathyaprakash & Schutz 2009; Colpi & Sesana 2017). At the lowest frequencies around 10^{-9}–10^{-7} Hz, Pulsar-Timing Array experiments (PTA) aim at detecting the GW signal from a population of MBHBs with masses around 10^8–$10^{10}\,M_\odot$, thousands to millions of years prior to coalescence (Sazhin 1978; Foster & Backer 1990; Rajagopal & Romani 1995; Wyithe & Loeb 2003; Jaffe & Backer 2003a; Enoki et al. 2004; Sesana et al. 2004). Although PTA experiments are also sensitive to GWs from single MBHBs, the most likely signal to be detected first is a stochastic gravitational background (GWB) produced by the

★ E-mail: david.izquierdovillalba@unimib.it

incoherent superposition of GWs from the cosmic population of inspiralling MBHBs out to $z \sim 1$ (Rosado, Sesana & Gair 2015). To detect such signal, PTA experiments search for spatial correlated fluctuations in the times of arrival of radio pulses from a network of millisecond pulsars in the Milky Way. Currently, three main PTA experiments are taking data: the *European Pulsar Timing Array* (EPTA; Kramer & Champion 2013; Desvignes et al. 2016), the *North American Nanohertz Observatory for Gravitational Waves* (NANOGrav; McLaughlin 2013; Arzoumanian et al. 2015), and *Parkes Pulsar Timing Array* (PPTA; Manchester et al. 2013; Reardon et al. 2016) projects. The three collaborations share data under the aegis of the International PTA (IPTA; Hobbs et al. 2010; Perera et al. 2019). The final goal is to construct a global PTA with all the data collected around the world, including those provided by recently formed PTAs – such as the Indian PTA (InPTA; Susobhanan et al. 2021) and the Chinese PTA (CPTA; Lee 2016) – and by cutting-edge new timing instruments like MeerKAT (Bailes et al. 2016). In the last decade, EPTA, NANOGrav, PPTA, and IPTA have been collecting data of ever improving quality, publishing a number of upper limits to the amplitude of the GWB $A_{\rm yr^{-1}} \lesssim (2\text{–}3) \times 10^{-15}$ at $1\,{\rm yr}^{-1}$ (Lentati et al. 2015; Shannon et al. 2015; Verbiest et al. 2016; Arzoumanian et al. 2018). Interestingly, the most recent results of NANOGrav (12.5-yr data set), PPTA (second data release, DR2) and EPTA (DR2) have pointed out the existence of a stochastic process with median amplitude of $A_{\rm yr^{-1}} \sim (1.9\text{–}2.95) \times 10^{-15}$ (Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021). However, the lack of significant evidence of the quadrupolar correlations in such detected signals makes difficult to claim a GWB detection.

From a theoretical point of view, several works aim at predicting the expected stochastic GWB at n-Hz frequencies. For instance, Jaffe & Backer (2003b) reported an amplitude of $A_{\rm yr^{-1}} \sim 10^{-16}$ by linking the observed merger rate of massive galaxies with some analytical prescriptions for MBH binary evolution. However, Wyithe & Loeb (2003) showed that $A_{\rm yr^{-1}}$ could increase up to $\sim 10^{-15}$ if the galaxy merger rate is computed from the extended Press and Schechter theory (PS; Press & Schechter 1974). Such discrepancies were principally due to the different analytical recipes used to treat the DM halo and black hole physics, which, in turn, reflected the lack of knowledge about how haloes and MBH binaries co-evolve with cosmic time. Indeed, the large variance caused by such effect was noticed by Sesana, Vecchio & Colacino (2008), who carried out a systematic study on the GW stochastic background predicted by a wide variety of semi-analytical models (SAMs) based on the PS halo mass function. The authors concluded that taking into account the uncertainties of all these models, the expected GWB amplitude detected by PTA could expand between $A_{\rm yr^{-1}} \sim 2.4 \times 10^{-16}$ and $\sim 3.8 \times 10^{-15}$. To improve the statistics of the PS haloes and to avoid the overproduction of low-z bright quasar seen in PS-based models (e.g. Marulli et al. 2006), a number of works used merger trees extracted form cosmological N-body simulation. Among them, we cite Sesana, Vecchio & Volonteri (2009), who explored the PTA predictions using the catalogue of merging galaxies extracted from Bertone, De Lucia & Thomas (2007) SAM applied on the `Millennium` DM merger trees (Springel 2005). By associating to each merging galaxy a central MBH according to some observational prescription, the authors reported $4 \times 10^{-16} < A_{\rm yr^{-1}} < 2 \times 10^{-15}$. Besides, Sesana et al. (2009) concluded that depending on the model used for placing MBHs, individual signals from MBHBs could be detected in PTA data. However, these types of events are likely to be rare. Similar work was performed by Roebber et al. (2016) using the N-body simulations `Dark Sky` and `MultiDark` (Riebe et al. 2011; Skillman et al. 2014): Placing galaxies and MBHs inside DM haloes through scaling relations and leaving aside a detailed modelling of binary dynamics and associated delay after the halo–halo merge, the authors found a typical value of $A_{\rm yr^{-1}} \sim 6 \times 10^{-16}$. Another class of models directly exploits observations of galaxy pairs to infer a galaxy and MBHB merger rate, which is then used to construct a stochastic GWBs. Such models were extensively investigated by Sesana (2013), Ravi et al. (2015), and Sesana et al. (2016), yielding $3 \times 10^{-16} < A_{\rm yr^{-1}} < 2 \times 10^{-15}$, due to uncertainties in defining galaxy pairs, estimating merger time-scales and connecting MBHs to their hosts via a bulge–MBH mass relations (Kormendy & Ho 2013; Shankar et al. 2016).

Even though all these models were already providing strong constraints on the GWB at n-Hz frequencies, they relied on uncertain observations and/or empirical relations to place galaxies and MBHs in the DM merger trees. Crucially, they missed a self-consistent treatment of galaxy evolution and of how MBHs and MBHB form and evolve inside galaxies. To improve these limitations, Dvorkin & Barausse (2017) and Bonetti et al. (2018a) based their GWB predictions on the SAM of Barausse (2012). This model, based on PS merger trees, had the advantage of including a detailed modelling for the cosmological evolution of galaxies and MBHs, and it further refined to include different prescriptions for the MBH binary evolution. On one side, Dvorkin & Barausse (2017) explored the GWB amplitude within the PTA band in the worst possible scenario, i.e. if all MBHBs are not able to merge and they are stalled at ∼pc scales (i.e. the so-called *final-parsec problem*; Milosavljević & Merritt 2001). Their results showed that even in this pessimistic scenario, a GW signal should remain in the PTA band ($A_{\rm yr^{-1}} \sim 10^{-16}$). On the other hand, Bonetti et al. (2018a) performed a similar study but extending the treatment of MBH binaries and including a refined model of triple MBH interactions as a plausible mechanism for avoiding the stalling of MBHBs. The authors reported $A_{\rm yr^{-1}} \sim 10^{-15}$ and highlighted that triple interactions between an MBHB and an MBH orbiting around the binary or impinging on it, play an important role in the final GWB amplitude, avoiding the reduction of the signal as a consequence of the stalling binaries. Thanks to the fast development of cosmological hydrodynamical simulations able to follow the assembly of galaxies down to relatively small scales in large cosmological volumes, recent works have also drawn predictions for $A_{\rm yr^{-1}}$ by taking advantage of the galaxy properties provided by these simulations. Kelley, Blecha & Hernquist (2017a) used the galaxy population of the `Illustris` simulation (Vogelsberger et al. 2014a,b) to construct a comprehensive modelling for tracking the different evolutionary stages of MBH binaries. With such **a** model, Kelley et al. (2017a) reported $A_{\rm yr^{-1}} \sim 7 \times 10^{-16}$ with most of the signal coming from very massive binaries ($\sim 10^9\,{\rm M}_\odot$) merging at low-z ($z < 3$). This work was extended by Siwek, Kelley & Hernquist (2020), who explored the repercussion of gas accretion in MBH binaries on the GWB. Their results showed that if the growth of the secondary MBH is favoured, the GWB level could reach up to $A_{\rm yr^{-1}} \sim 10^{-15}$. In contrast, in the case in which the secondary MBH growth is halted, the GWB dropped down to $A_{\rm yr^{-1}} \sim 3 \times 10^{-16}$.

A fact worth noticing is that GWB amplitudes up to $A_{\rm yr^{-1}} \approx 4 \times 10^{-15}$ can be found in the literature. However, models that self-consistently evolve galaxies and MBHs and that reproduce the MBH mass and quasar luminosity functions hardly get a GWB level much in excess of $A_{\rm yr^{-1}} \approx 1 \times 10^{-15}$ (Kelley et al. 2017b; Bonetti et al. 2018a). This is particularly interesting in light of the recent results of the NANOGrav (12.5-yr data set), PPTA (DR2) and EPTA (DR2) collaboration that reported strong evidences of a stochastic process with $A_{\rm yr^{-1}}$ spanning, respectively, between 1.37×10^{-15}–2.67×10^{-15}, 1.9×10^{-15}–2.6×10^{-15}, and 2.23×10^{-15}–3.8×10^{-15}

(Arzoumanian et al. 2020; Chen et al. 2021; Goncharov et al. 2021). Note that the signal seen by NANOGrav, PPTA and EPTA did not display significant evidence of the quadrupolar correlations needed to claim detection of a GWB. Never the less, it is fundamental to explore what theoretical models can tell us about such a large signal level. Even more, precision in the measurement of the GWB amplitude could be used as a new tool to improve our knowledge about the co-evolution of MBHs and galaxies, rule out theoretical models of MBH binary evolution, and test our current treatment of galaxy formation, whose detailed modelling is still a challenge.

Motivated by this, in this work, we explore the evolution of MBH binaries in the context in galaxy formation models. For that, we introduce a model of MBHB formation and evolution embedded inside the L-Galaxies SAM in the version of Izquierdo-Villalba et al. (2019, 2020). Specifically, unlike many other SAMs in the literature, the model introduces recipes for the MBH dynamics in the host galaxy, as the MBHB coalescence is not instantaneous. The MBHs need to reach sub-pc scales for GWs to drive the evolution and enter the PTA bandwidth. Thus, stellar and gas dynamical torques acting on galactic scales lead to delays in the computation of the binary merger time-scale compared to the time-scale of the colliding galaxies. All these processes have been included in a self-consistent manner inside the cosmological evolution of galaxies and black holes tracked by L-Galaxies. We have applied the new model on the Millennium DM merger trees (Springel 2005) whose box-size and mass resolution had offered us the capability of drawing predictions for GW emission in the PTA band. To our knowledge, this work is the first to include current GWB measurements as an extra constraint to calibrate the evolution of MBHs within the context of galaxy formation models. In particular, we add the GWB to the standard constraints provided by the quasars luminosity function (QLF) and mass function of MBHs (BHMF) in the local Universe (Marconi et al. 2004; Hopkins et al. 2007; Shankar, Weinberg & Miralda-Escudé 2009; Shen et al. 2020). In this way, we are able to explore, for the first time, how feasible is for these galaxy formation models (and, in particular, our version of L-Galaxies) to jointly reach the current measurements of the GWB while reproducing the well-constrained QLF and BHMF.

This paper is organized as follows: In Section 2, we describe the main characteristics of L-Galaxies and Millennium simulations. In Section 3, we present the model that traces the formation and evolution of MBHBs. In Section 4, we present our results, focusing on the GW signal in the PTA frequency band and the difficulties of the model to produce large GW amplitudes without mismatching other MBH constrains such as MBH mass function. A Lambda cold dark matter (ΛCDM) cosmology with parameters $\Omega_m = 0.315$, $\Omega_\Lambda = 0.685$, $\Omega_b = 0.045$, $\sigma_8 = 0.9$, and $H_0 = 67.3\,km\,s^{-1}\,Mpc^{-1}$ is adopted throughout this paper (Planck Collaboration XVI 2014).

2 GALAXY FORMATION MODEL

In the following sections, we briefly overview the main physics included in the L-Galaxies SAM. L-Galaxies is a code that tracks the time evolution of gas, stars, and MBHs within their host DM subhaloes[1] through a series of differential equations and analytic prescriptions. The version of the model used here is the Henriques et al. (2015) but with the modifications in the bulge and black hole physics presented in Izquierdo-Villalba et al. (2019, 2020).

[1]In this work, we define subhaloes as locally overdense, self-bound particle groups formed inside the DM haloes.

2.1 DM merger trees

DM merger trees are the backbone of any SAM. In this paper, we use the trees extracted from the Millennium *N*-body simulation (hereafter MS; Springel 2005). MS follows the cosmological evolution of 2160^3 DM particles with a mass of $8.6 \times 10^8\,M_\odot\,h^{-1}$ within a periodic cube of 500 $Mpc\,h^{-1}$ on a side. Even though MS was run by using WMAP1 and 2dFGRS cosmology, the version of L-Galaxies in this work is tuned on a re-scaled versions of the MS simulation (Angulo & White 2010) to match the cosmological parameters obtained by Planck first-year data release (Planck Collaboration XVI 2014).

All the particle information of MS is stored at 63 different epochs or *snapshots*. At every snapshot DM haloes and subhaloes are extracted using a friend-of-friend (FOF) group-finder and SUBFIND algorithm (Springel et al. 2001), respectively. By applying L-HALOTREE (Springel 2005), all halo and subhalo structures are arranged in merger trees to follow the evolutionary path of any DM (sub)halo in the simulations. We highlight that L-Galaxies is based on the subhalo population instead of the halo one. This enables L-Galaxies to build up a more realistic galaxy population, making more reasonable predictions on the galaxy merger rate and clustering. Never the less, the time resolution given by the 63 snapshots is not enough to properly trace the baryonic physics. Thus, the SAM does an internal time discretization between two consecutive snapshots with approximately ∼5–20 Myr of time resolution. These extra-temporal subdivisions of L-Galaxies are called *sub-steps*.

2.2 Baryonic physics

L-Galaxies follows the standard scenario of structure formation, by assuming that when a subhalo virializes, part of the diffuse baryonic gas present in its surroundings is trapped and collapses within it. Baryons are deposited in the subhalo in the form of a hot gas atmosphere. Within the cooling time-scale, this gas gradually migrates towards the centre of the subhalo, forming a disc-like structure, called cold-gas disc. When the disc is large enough, episodes of star formation are triggered, leading to the assembly of the stellar disc. L-Galaxies self-regulates the formation of stars by including feedback both from a central AGN and supernovae. Galaxies are able to form a overdensity of stars in the nuclear region (i.e. the so-called bulge) via mergers and disc instabilities (DIs). According to the baryonic merger ratio of the two interacting galaxies, the remnant can be transformed into an elliptical galaxy, or can preserve the stellar disc developing a galactic bulge by incorporating the whole stellar component of the smaller progenitor. In the model used here, we introduce the concept of *smooth accretion*, which occurs when the less massive progenitor is completely absorbed by the stellar disc of the central galaxy (Izquierdo-Villalba et al. 2019). Alternatively, DIs in massive discs can change the stellar distribution, leading to the formation of a central ellipsoidal component, typically referred to as bar or pseudo-bulge.

2.3 Black hole physics: growth and spin

Each newly resolved subhalo (independently of redshift and halo properties) is seeded with a black hole of $10^4\,M_\odot$ whose spin has a modulus $|a|$ randomly selected between $0 < |a| < 0.998$. The choice of the initial seed mass is conservative, given the minimum mass of new resolved subhaloes in the MS ($\sim 10^{10}\,M_\odot$). In future works, we will explore the model predictions for MBHs using the refined seeding procedure presented in Spinoso et al. (in preparation). Once

the black hole seed is placed in its host galaxy, it can grow through three different channels: *cold gas accretion*, *hot gas accretion*, and *mergers* with other black holes. Specifically, the first channel is the main driver of the black hole growth and is triggered by both galaxy mergers and DI events. After a galaxy merger, we assume that the fraction of cold gas accreted by the nuclear black hole is

$$\Delta M_{\rm BH}^{\rm gas} = f_{\rm BH}^{\rm merger}(1+z_{\rm merger})^{5/2}\frac{m_{\rm R}}{1+(V_{\rm BH}/V_{200})^2}\,M_{\rm gas}, \quad (1)$$

where $m_{\rm R} = M_{\rm satellite}^{\rm baryon}/M_{\rm central}^{\rm baryon} \leq 1$ is the baryonic ratio of the two interacting galaxies, V_{200} is the virial velocity of the host DM subhalo, $z_{\rm merger}$ is the redshift of the galaxy merger, $M_{\rm gas}$ is the cold gas mass of the galaxy, and $V_{\rm BH}$, $f_{\rm BH}^{\rm merger}$ are two adjustable parameters set to 280 km s^{-1} and 0.025, respectively. In the presence of a DI, the black hole accretes an amount of cold gas proportional to the mass of stars that trigger the stellar DI, $\Delta M_{\rm stars}^{\rm DI}$ [2]:

$$\Delta M_{\rm BH}^{\rm gas} = f_{\rm BH}^{\rm DI}(1+z_{\rm DI})^{5/2}\frac{\Delta M_{\rm stars}^{\rm DI}}{1+(V_{\rm BH}/V_{200})^2}, \quad (2)$$

where $z_{\rm DI}$ is the redshift in which the DI takes place, and $f_{\rm BH}^{\rm DI}$ is a free parameter that takes into account the gas accretion efficiency, set to 0.0015. We highlight that the redshift dependence of equations (1) and (2) has been modified with respect to Izquierdo-Villalba et al. (2020) to improve the match between the observed and the predicted black hole mass function and bulge-MBH correlations at $z = 0$.

After a galaxy merger or a DI, the cold gas available for accretion is assumed to settle in a reservoir around the black hole, $M_{\rm Res}$. Instead of an instantaneous gas consumption, the model considers that the gas reservoir is progressively consumed trough a Eddington-limited growth phase, followed by a second phase of low accretion rates (Hopkins et al. 2005, 2006b; Marulli et al. 2006; Bonoli et al. 2009). We refer the reader to Izquierdo-Villalba et al. (2020) for further details.

During any of the events that make the MBH grow, the code tracks the evolution of the black hole spin in a self-consistent way. During gas accretion events, the model uses the approach presented in Dotti et al. (2013) and Sesana et al. (2014), which links the number of accretion events that spin-up or spin-down the MBH with the degree of coherent motion in the bulge. In particular, the model assumes that DIs increase the coherence of the bulge kinematics. On the other hand, mergers bring disorder to the bulge dynamics. After a MBH coalescence the final spin is determined by the expression of Barausse & Rezzolla (2009), where a distinction between wet and dry mergers is done to compute the alignment/anti-alignment between the two MBHs. For further details on the implementation in the SAM, we refer the reader to Izquierdo-Villalba et al. (2020). We highlight that in this work, we do not include the gravitational recoils after coalescence, as presented in Izquierdo-Villalba et al. (2020). In a future work, we will explore what is the effect of recoils on the population of MBHBs.

Finally, we highlight that all the parameters used in the SAM (including the ones of equations 1 and 2) have been chosen to reproduce many observed galaxy and MBH properties. Among them, we can highlight the stellar mass function, the fraction of passive galaxies, quasar luminosity function, the $z = 0$ black hole mass function, or the $z = 0$ correlation between bulge and black hole mass (we refer to Henriques et al. 2015 and Izquierdo-Villalba et al. 2020 for the specific comparisons).

3 THE POPULATION OF MASSIVE BINARY BLACK HOLES

In this section, we describe the physics included in `L-Galaxies` to follow the formation and coalescence of MBHBs. Following Begelman, Blandford & Rees (1980), we divide the evolutionary pathway of MBHBs into three stages. The first one is described in Section 3.1 and consists of a *pairing* phase in which, after the galaxy merger, the dynamical friction exerted by the stars drives the MBH of the satellite galaxy toward the nucleus of the remnant galaxy where it binds with the central MBH. This occurs when the amount of stars enclosed within the binary orbit is comparable to the mass of the lighter MBH of the binary. Then, a *hardening* phase takes place in which the orbital semi-major axis of the binary shrinks due to three-body interactions with single stars (the slingshots mechanism) and/or interaction with a massive gaseous circumbinary disc (Colpi 2014). Finally, a *GW inspiral* phase drives the binary to coalescence. We discuss the implementation of the two last phases in Section 3.2.

3.1 The pairing phase of MBHs

The first phase that anticipates the formation of a binary system at the centre of the post-merger galaxy consists in reducing the MBH separation from ∼kpc to ∼pc through dynamical friction. In this work, to estimate the time spent by a black hole in the pairing phase, we use the expression (Binney & Tremaine 2008)

$$t_{\rm dyn}^{\rm BH} = 19\,f(\varepsilon)\left(\frac{r_0}{5\,{\rm kpc}}\right)^2\left(\frac{\sigma}{200\,{\rm km\,s^{-1}}}\right)\left(\frac{10^8\,{\rm M_\odot}}{M_{\rm BH}}\right)\frac{1}{\Lambda}\,[{\rm Gyr}], \quad (3)$$

where $f(\varepsilon)$ is a function that depends on the orbital circularity of the black hole ε (Colpi, Mayer & Governato 1999), r_0 is the initial position of the black hole deposited by the satellite galaxy after the merger, σ is the velocity dispersion of the remnant galaxy ($\sigma^2 = 0.25{\rm G}M_{\rm stellar}/R_{\rm gal}$),[3] $M_{\rm BH}$ is the mass of the black hole, and $\Lambda = \ln(1 + M_{\rm stellar}/M_{\rm BH})$ is the Coulomb logarithm (Mo, van den Bosch & White 2010).

The value of r_0 in equation (3) is the position where the satellite galaxy has lost a fraction $F_{\rm strp}$ of its total mass by tidal stripping. Such position is determined by solving numerically the expression (King 1962; Taylor & Babul 2001)

$$\frac{{\rm d}^2\Phi(r)}{{\rm d}r^2} = \omega^2 - \frac{{\rm G\,M_{sat}}(<R)}{R^3}, \quad (4)$$

where the variable r is the radial position of the satellite galaxy within the DM subhalo, ω is its instantaneous orbital angular velocity, and Φ the potential of the hosting subhalo. Finally, R and $M_{\rm sat}(<R)$ represent the radius and mass at which the satellite galaxy contains $(1-F_{\rm strp})$ of its total baryonic mass. While the value of $M_{\rm sat}(<R)$ is computed assuming exponential disc and Sérsic bulge profiles (Sersic

[2] DIs are accounted for by `L-Galaxies` using the Efstathiou, Lake & Negroponte (1982) criterion. Based on that prescription, the amount of matter that triggers a DI event is set to

$$\Delta M_{\rm stars}^{\rm DI} = M_{\star,{\rm d}} - \left(V_{\rm max}^2 R_{\star,{\rm d}}/G\epsilon^2\right) > 0,$$

where ϵ is a free parameter set to 1.5, $V_{\rm max}$ is the maximum circular velocity of the host DM, and $R_{\star,{\rm d}}$ and $M_{\star,{\rm d}}$ are the length and stellar mass of the stellar disc, respectively.

[3] $R_{\rm gal}$ refers to the effective radius of the galaxy, computed as the mass weighted average of the galaxy bulge and stellar disc radius. In Izquierdo-Villalba et al. (2019), it was shown that $R_{\rm gal}$ values predicted by `L-Galaxies` are compatible with current observations.

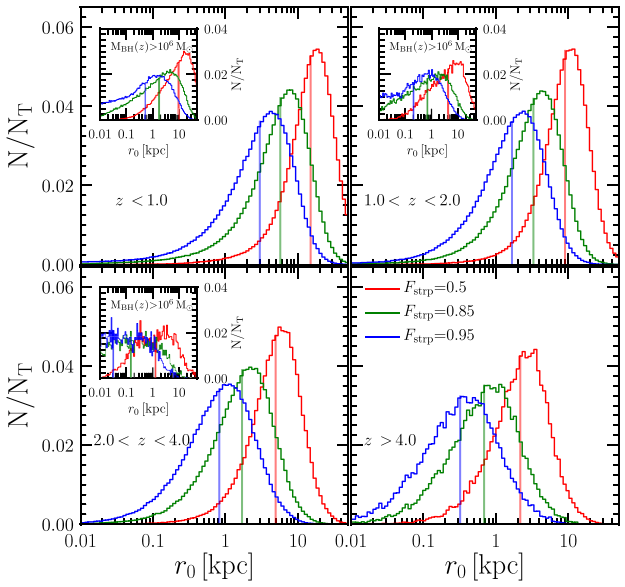

Figure 1. Distribution of r_0 for three different values of $F_{\rm strp}$ representing the mass lost by the secondary due to tidal stripping by the primary galaxy: 0.5 (red), 0.85 (green), and 0.95 (blue). Solid vertical lines represent the median value for each distribution. Each panel displays a different redshift bin. The inner panels show the same but only for satellite galaxies that deposit an MBH of mass $>10^6$ M$_\odot$. We do not show the cases at $z > 4$, given the small number of satellite galaxies with $>10^6$ M$_\odot$ MBHs at that high z.

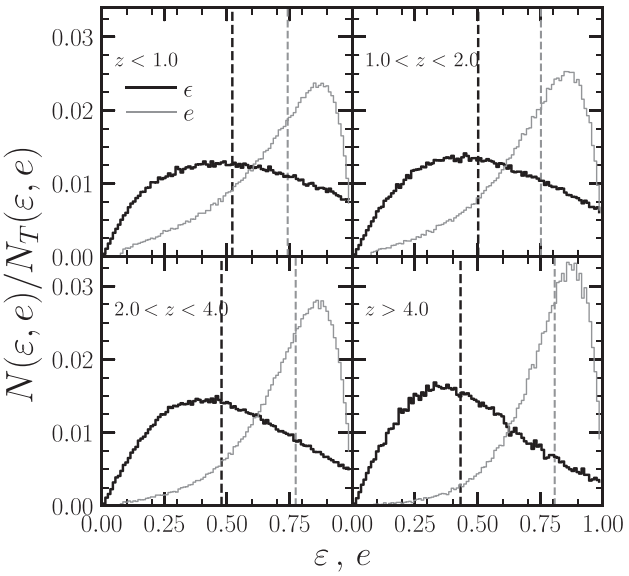

Figure 2. Distribution of circularity (ε, thick black line) and eccentricity (e, thin grey line) of the satellite black holes at the moment in which they are deposited at r_0. Different panels represent different redshifts bins and vertical lines represent the median of the distribution.

1968), the subhalo potential is modeled as a *Navarro–Frenk–White* (NFW; Navarro, Frenk & White 1996).[4] Given the limitations of `L-Galaxies` to provide accurate positions of satellite galaxies that had lost their DM subhalo, we evaluate the quantities of equation (3) at the instant at which the DM subhalo associated with the satellite galaxy merges with the one associated with the central galaxy. From this moment, the DM host of the satellite is not resolved anymore by the DM simulation.

In Fig. 1, we present the distribution of r_0 for three different values of $F_{\rm strp}$ (0.5, 0.85, and 0.95). As we can see, the larger the $F_{\rm strp}$, the smaller is r_0. Moreover, regardless of $F_{\rm strp}$, there is a redshift evolution in the r_0 values. In particular, the smaller is the redshift, the larger is the typical r_0. This is a consequence of the increase of DM halo mass and its concentration towards low redshifts (Dutton & Macciò 2014), which causes the halo potential to be more efficient in disrupting the satellite galaxy. To check if the r_0 distribution changes for the most massive MBHs, in the inner plots of Fig. 1, we present the values of r_0 only for satellite galaxies that deposit a $>10^6$ M$_\odot$ MBH. As shown, these galaxies follow the general trend of the large $F_{\rm strp}$ values being associated with small r_0 values. Never the less, regardless of $F_{\rm strp}$, they have a median r_0 smaller than the general population. This deviation is caused because the former population have stellar masses $\sim$1 dex larger: $M_{\rm stellar} \sim 10^{9.5}$ M$_\odot$ versus $M_{\rm stellar} \sim 10^{8.7}$ M$_\odot$ of the general satellite population. This mass difference causes that satellite galaxies hosting $> 10^6$ M$_\odot$ MBHs take more time before being stripped, having more chances to deposit the MBH at low r_0 values. In this work, we decided to use $F_{\rm strp} = 0.85$. Even though this choice is somewhat arbitrary, we selected such a high threshold to be sure that most of the stellar component around the satellite MBH is already tidally removed by the merging process. Thus, the dynamics of the MBH can be progressively considered as the one of a *naked MBH* moving in the stellar background of the remnant galaxy.

As shown in equation (3), the dynamical friction time-scale depends on the circularity of the MBH orbit. Following Lacey & Cole (1993), we adopt $f(\varepsilon) \sim \varepsilon^{0.78}$ (see other methods as Colpi et al. 1999; Boylan-Kolchin, Ma & Quataert 2008). Here we assume that that the MBH orbital circularity is inherited form the one of the satellite galaxy. Specifically, galaxy orbital circularities are computed following Scannapieco et al. (2009) (see also Abadi et al. 2003) defining ε as

$$\varepsilon = \frac{j}{j_{\rm c}}, \tag{5}$$

where j is the angular momentum per unit of mass of the satellite galaxy at a distance r from the halo centre and $j_{\rm c}$ is the angular momentum expected for a circular orbit at the same r, i.e. $j_{\rm c} = r\, v_{\rm c}(r) = r\sqrt{GM_{\rm h}(<r)/r}$, where $M_{\rm h}(<r)$ is the halo mass enclosed within r, computed assuming an NFW profile. As we did before, equation (5) is computed as soon as the satellite galaxy looses its host DM subhalo. In Fig. 2, we show the orbital circularity of the MBHs in the `Millennium` DM merger trees. As we can see, ε has a moderate evolution with redshift. The peak around $\varepsilon \approx 0.35$ gets progressively smeared out, with larger circularities becoming more common at lower redshift. We also show for completeness the distribution of orbital eccentricities.[5] As we move to lower redshifts, the distribution tends to develop a substantial tail at low values,

[4]Given that the `Millennium` merger trees catalogues do not contain the subhalo concentration, we use the fits of Dutton & Macciò (2014) to obtain their concentration at any redshift and mass.

[5]This value has been computed as $e = (r_+ - r_-)/(r_+ + r_-)$, with r_+ and r_- being the apo- and pericentre of the orbit. Such quantities are the roots of $\left(1/r^2\right) + \left(2\left[\Phi(r) - E\right]/L^2\right) = 0$. The values of E and L are, respectively, the energy and angular momentum per unit mass in a spherical potential (Φ, in our case the NFW potential).

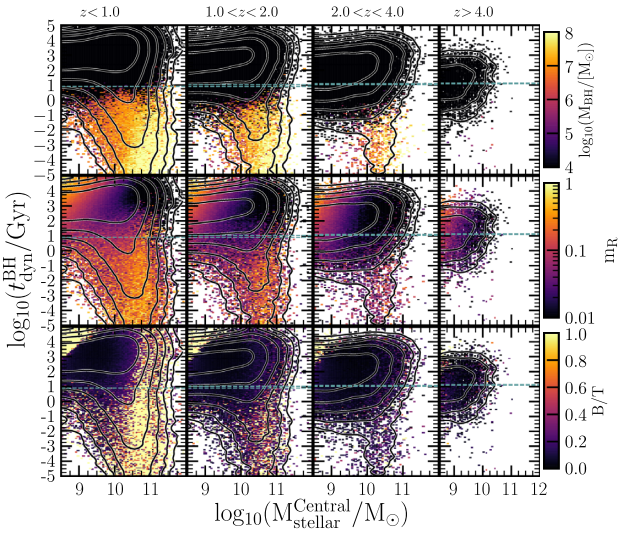

Figure 3. $M_{\rm stellar}^{\rm Central}$–$t_{\rm dyn}^{\rm BH}$ plane at four different redshift bins ($z < 1$, $1 < z < 2$, $2 < z < 4$, and $z > 4$). The first row encodes the median black hole mass of the MBH in the pairing phase ($M_{\rm BH}$) at a given bin of $M_{\rm stellar}^{\rm Central}$ and $t_{\rm dyn}$. In the second row, the colour represents the baryonic galaxy merger ratio, $m_{\rm R}$. The colour map of the third panel encodes the bulge-to-total ratio (B/T) of the hosting galaxy. In all the panels, the blue lines represent the Hubble time ($t_{\rm H}$) at a given redshift bin. In all the panels, the white lines represent the contours where a same number of MBHs is enclosed it.

although it maintains a maximum at $e \approx 0.8$, in agreement with Tormen (1997).

Fig. 3 carries information on key quantities in the plane $M_{\rm stellar}^{\rm Central}$–$t_{\rm dyn}^{\rm BH}$, where the former is the stellar content of the post-merger galaxy. We show the results for $M_{\rm stellar}^{\rm Central} > 10^{8.5}{\rm M}_{\odot}$ corresponding to the range above which the results are not significantly affected by resolution of the underlying `Millennium` DM simulation. In each panel, the distribution has been colour coded by the mass of the satellite MBH in the pairing phase ($M_{\rm BH}$), the baryonic merger ratio of the two interacting galaxies ($m_{\rm R}$) and the *bulge-to-total* ratio of the remnant galaxy (B/T). At $z > 4$, there is a significant fraction of satellite MBHs (∼47 per cent) that would reach the centre of the galaxy within the Hubble time ($t_{\rm H}$). This is principally caused by the fact that at high-z, DM subhaloes are smaller and galaxies more compact. Due to the combination of these two facts, satellite galaxies are less affected by strong tidal effects and thus capable deposit the MBH at closer distances from the nucleus of the central galaxy (see r_0 of Fig. 1). Interestingly, most of these MBHs are close to the seed mass ($10^4\,{\rm M}_{\odot}$), which is a direct consequence of the rough seeding procedure used in this work. On the other hand, at lower redshifts the situation changes and the number of MBHs with $t_{\rm dyn}^{\rm BH} > t_{\rm H}$ is the predominant (>80 per cent of the cases). Even though the fraction of MBHs that merge in a Hubble time is decreasing, essentially all systems with a MBH larger than a $10^6\,{\rm M}_{\odot}$ do so. Indeed, all the non-merging systems involve small MBHs that are leftovers from the seeding procedure. In future work, we will use the model of Spinoso et al. (in preparation) to explore the effect of seeding in the population of MBHBs.

Regarding the merger ratio, events with $t_{\rm dyn}^{\rm BH} < t_{\rm H}$ display $m_{\rm R} > 0.1$ at all redshifts. At stellar masses $>10^{10}\,{\rm M}_{\odot}$ we find cases with $t_{\rm dyn}^{\rm BH} > t_{\rm H}$ characterized by very low $m_{\rm R}$ ($\lesssim$0.01). We checked that these events corresponds to minor mergers between massive galaxies and small galaxy companions ($<10^9\,{\rm M}_{\odot}$) whose host nuclear MBH rarely exceeds $10^5\,{\rm M}_{\odot}$. At $M_{\rm stellar}^{\rm Central} < 10^{10}\,{\rm M}_{\odot}$, it is less common to have events with small merger ratios, especially at $z < 2$. This is a natural consequence of the `Millennium` resolution as the minimum resolved stellar mass of satellite galaxies, $\sim 10^8\,{\rm M}_{\odot}$, is comparable with the mass of the central galaxy for $M_{\rm stellar}^{\rm Central} < 10^{10}\,{\rm M}_{\odot}$ (see fig. B2 of Izquierdo-Villalba et al. 2019). Despite the large $m_{\rm R}$ of these events, $t_{\rm dyn}^{\rm BH}$ values are on average relatively large. This is caused by both the small mass of the black holes ($< 10^5\,{\rm M}_{\odot}$) and the large r_0 (∼10 kpc) characterizing these events. Finally, the plane $M_{\rm stellar}^{\rm Central}$–$t_{\rm dyn}^{\rm BH}$ seems to display a correlation with the galaxy morphology. In particular, the larger is the B/T the smaller is $t_{\rm dyn}^{\rm BH}$. This effect is particularly evident at $z < 2$, where elliptical galaxies (B/T > 0.7) host pairing black holes with lower $t_{\rm dyn}^{\rm BH}$.

3.2 Hardening and GW phase

As soon as the pairing phase ends,[6] we assume that the MBHs form a hard binary with the central one. From hereafter, we tag as *primary* black hole (with mass $M_{\rm BH,1}$) the most MBH in the system whereas the less massive one is refereed as *secondary* black hole (with mass $M_{\rm BH,2}$). The initial semi-major axis of the binary, $a_{\rm BH}$, is set to the scale in which $M_{\rm Bulge}(< a_{\rm BH}) = 2M_{\rm BH,2}$, where $M_{\rm Bulge}(< a_{\rm BH})$ is the mass in stars of the hosting bulge within $a_{\rm BH}$. In this work we assume that the evolution of the binary system depends on the type of environment in which is hosted. In particular, following Antonini, Barausse & Silk (2015), we distinguish between two different type of environments that drive the two MBHs to final coalescence: mergers in *gas-rich* and *gas-poor* environments.

Mergers in gas-rich environments require the binary to be surrounded by a gas reservoir with a mass larger than the mass of the binary (i.e. $M_{\rm Res} > M_{\rm Bin}$; Antonini et al. 2015). In this case, the shrinking of the binary separation and the subsequent final coalescence is driven by the interaction with a massive circumbinary disc and GW emission. This scenario is supported by the results of the hydrodynamical simulations of Escala et al. (2004, 2005), Dotti et al. (2007), and Cuadra et al. (2009), which showed that dense gaseous circumbinary discs are effective in shrinking MBHBs, promoting their coalescence in less than $\lesssim 10^7$ yr (see also the work of Armitage & Natarajan 2002; Kocsis, Haiman & Loeb 2012). Given such effectiveness of the circumbinary gas discs in driving the MBHB to the final coalescence, we neglect the stellar hardening effect. In this work, we follow the results of Dotti, Merloni & Montuori (2015) (see also Bonetti et al. 2019), assuming that the evolution of the binary semi-major axis can be inferred from

$$\frac{{\rm d}a_{\rm BH}}{{\rm d}t} = \left(\frac{{\rm d}a_{\rm BH}}{{\rm d}t}\right)_{\rm Gas} + \left(\frac{{\rm d}a_{\rm BH}}{{\rm d}t}\right)_{\rm GW} = -\frac{2\dot{M}_{\rm Bin}}{\mu}\sqrt{\frac{\delta}{1-e_{\rm BH}^2}}\,a_{\rm BH} - \frac{64G^3(M_{\rm BH_1}+M_{\rm BH_2})^3F(e_{\rm BH})}{5c^5(1+q)^2a_{\rm BH}^3}, \quad (6)$$

where the first and second term take into account the gas hardening and GW emission, respectively. Regarding the variables, G is the gravitational constant, c is the light speed, $\delta = (1+q)(1+e_{\rm BH})$, $q = M_{\rm BH,2}/M_{\rm BH,1}$, $\dot{M}_{\rm Bin}$ is the sum of the accretion rate of both MBHs in the binary, and μ is the reduced mass of the binary. Finally,

[6] We assume that the pairing phase ends when $(t_{\rm merge}^{\rm Gal}$–$t_{\rm now})$–$t_{\rm dyn}^{\rm BH} < 0$. $t_{\rm merge}^{\rm Gal}$ correspond to the lookback time at which the galaxy merger takes place and and $t_{\rm now}$ is the lookback time of the simulation.

$F(e)$ is a function that depends on the binary eccentricity (Peters & Mathews 1963),

$$F(e_{\rm BH}) = (1 - e_{\rm BH})^{-7/2} \left[1 + \left(\frac{73}{24}\right) e_{\rm BH}^2 + \left(\frac{37}{96}\right) e_{\rm BH}^4\right]. \quad (7)$$

Here, we assume a fixed initial value of $e_{\rm BH} = 0.6$ when the dynamics is gas-dominated and the binary is surrounded by a circumbinary disc (first term in equation 6). This value is motivated by the work of Roedig et al. (2011), who found that the binary eccentricity coasts to a constant value of ∼0.6. As soon as the GW emission (second term in equation 6) dominates the MBHB evolution, we track the eccentricity evolution as (Sesana, Haardt & Madau 2006)

$$\frac{{\rm d}e_{\rm BH}}{{\rm d}t} = -\frac{304}{15} \frac{G^3 q (M_{\rm BH_1} + M_{\rm BH_2})^3}{c^5 (1+q)^2 a_{\rm BH}^4 (1 - e_{\rm BH}^2)^{5/2}} \left(e_{\rm BH} + \frac{121}{304} e_{\rm BH}^3\right), \quad (8)$$

We highlight that if a binary system evolving in a gas-rich environment exhausts the gas reservoir before the final coalescence, we switch to the equations describing the evolution in gas-poor environments, which we now provide.

For mergers in gas-poor environments, we assume that the gas reservoir around the MBHs is smaller than the total mass of the binary (i.e. $M_{\rm Res} < M_{\rm Bin}$). In this case, the hardening is caused by the extraction of binary energy and angular momentum through three-body interactions with background stars that cross the binary orbit (Quinlan & Hernquist 1997; Sesana et al. 2006; Vasiliev, Antonini & Merritt 2014; Sesana & Khan 2015). As for the gas-rich case, the emission of GW starts to dominate when the hardening time becomes comparable to the GW time-scale. In particular, in these types of environments, the binary separation is tracked by integrating numerically the equation (Sesana & Khan 2015)

$$\begin{aligned}\frac{{\rm d}a_{\rm BH}}{{\rm d}t} &= \left(\frac{{\rm d}a_{\rm BH}}{{\rm d}t}\right)_{\rm Stars} + \left(\frac{{\rm d}a_{\rm BH}}{{\rm d}t}\right)_{\rm GW} \\ &= -\frac{G H \rho_{\rm inf}}{\sigma_{\rm inf}} a_{\rm BH}^2 - \frac{64 G^3 (M_{\rm BH_1} + M_{\rm BH_2})^3 F(e_{\rm BH})}{5 c^5 (1+q)^2 a_{\rm BH}^3}, \end{aligned} \quad (9)$$

where G is the gravitational constant, c is the light speed, and $H \approx 15$–20 is the hardening rate extracted from the tabulated values of Sesana et al. (2006). The values of $\rho_{\rm inf}$ and $\sigma_{\rm inf}$ correspond, respectively, to the density and velocity dispersion of stars at the MBHB sphere influence. For these types of environments, we assume that the binary systems start with an initial eccentricity randomly selected between $0 < e_{\rm BH} < 1$.[7] Besides, scattering experiments and numerical simulations in this type of environments indicate that the binary eccentricity is not constant during the hardening and GW phase but it changes through stellar encounters (Hills 1983; Mikkola & Valtonen 1992; Quinlan & Hernquist 1997; Sesana et al. 2006). In particular, the variation of the eccentricity of the MBHB can be expressed as

$$\begin{aligned}\frac{{\rm d}e_{\rm BH}}{{\rm d}t} &= \left(\frac{de_{\rm BH}}{{\rm d}t}\right)_{\rm Stars} + \left(\frac{de_{\rm BH}}{{\rm d}t}\right)_{\rm GW} \\ &= a_{\rm BH} \frac{G \rho_{\rm inf} H K}{\sigma_{\rm inf}} \\ &\quad - \frac{304}{15} \frac{G^3 q (M_{\rm BH_1} + M_{\rm BH_2})^3}{c^5 (1+q)^2 a_{\rm BH}^4 (1 - e_{\rm BH}^2)^{5/2}} \left(e_{\rm BH} + \frac{121}{304} e_{\rm BH}^3\right), \end{aligned} \quad (10)$$

where K is the eccentricity growth rate whose value is taken according to table 2 of Sesana et al. (2006).

The values $r_{\rm inf}$, $\rho_{\rm inf}$, and $\sigma_{\rm inf}$ of equations (9) and (10) were computed assuming a bulge mass profile. Unlike other works that use isothermal sphere or Dehnen profiles (see e.g Volonteri, Haardt & Madau 2003; Sesana 2010; Sesana & Khan 2015; Bonetti et al. 2018b; Volonteri et al. 2020), here we decided to use a Sérsic model. This choice is motivated by observational studies that found it to be a good approximation for fitting the bulge light distribution of different galaxies (Drory & Fisher 2007; Drory & Alvarez 2008; Gadotti 2009). The analytical expressions for the Sérsic model are taken from Prugniel & Simien (1997) (see also Terzić & Graham 2005):

$$\rho(r) = \rho_0 \left(\frac{r}{R_{\rm e}}\right)^{-p} {\rm e}^{-b(r/R_{\rm e})^{1/n}}, \quad (11)$$

$$\begin{aligned}\sigma^2(r) &= \frac{4\pi G \rho_0^2 {R_{\rm e}}^2 n^2 b^{2n(p-1)}}{\rho(r)} \\ &\quad \times \int_Z^\infty \mathcal{Z}^{-n(p+1)-1} {\rm e}^{-\mathcal{Z}} \gamma(n(3-p), \mathcal{Z}) {\rm d}\mathcal{Z}, \end{aligned} \quad (12)$$

where $R_{\rm e}$ is the bulge effective radius,[8] ρ_0 is the central bulge density, and n is its Sérsic index. This index correlates with the central concentration of the bulge, being the bulges with smaller n, the ones less centrally concentrated. Finally, the variable γ represents the incomplete gamma function, whereas Z, p, and b are three different quantities that depend on the bulge properties: $Z = b(r/R_{\rm e})^{1/n}$, $p = 1 - 0.6097 n^{-1} + 0.05563 n^{-2}$, and $b = 2n - 0.33 + 0.009\,876 n^{-1}$. This Sérsic model causes that smaller MBHs spend more time in the hardening phase than the most massive ones. To guide the reader, for an MBHB system with total mass $M_{\rm bin} = 10^9\,{\rm M}_\odot$, $q = 1$, and $e_{\rm BH} = 0.3$, the hardening time-scale is ∼0.2 Gyr. For the same system but with $M_{\rm bin} = 10^6\,{\rm M}_\odot$, the time increases up to 10 Gyr. For further details, we refer to Biava et al. (2019), where a detailed study of hardening time-scales in different bulge profiles was performed.

One of the disadvantages of `L-Galaxies` is that it does not compute Sérsic indexes, but only the mass assembled throughout different channels of growth: major and minor mergers assemble elliptical and classical bulges, whereas DIs prompt pseudo-bulges. To attach a Sérsic value to each galaxy, we compute the Sérsic index distribution of $z = 0$ pseudo-bulges, classical bulges, and elliptical galaxies using the observational data provided by Gadotti (2009). For each bulge type, we fit their distributions according to

$$f(n) = A \left(\frac{n}{n_0}\right) {\rm e}^{-(n/n_0)^\beta}, \quad (13)$$

where A, n_0, and β are free parameters. In Fig. 4, we show the fits for pseudo-bulges, classical bulges and elliptical galaxies. Table 1 contains the best fit for these parameters. As we can see, each bulge type follows a distinct distribution, and the larger differences are seen between pseudo-bulges and elliptical galaxies. Such difference have been reported in the last years, highlighting that the formation scenario of each bulge type might leave an imprint in the stellar dynamics and distribution (Kormendy 1983; Kormendy & Bender 1996; Drory & Fisher 2007; Drory & Alvarez 2008; Elmegreen, Bournaud & Elmegreen 2008; Gadotti 2009). Once the Sérsic index distribution is distributed, the way of assigning these values

[7] We have tested the model by assuming that the initial eccentricity of the hardening phase is inherited from the one computed in the pairing phase (see Section 3.1). We have found that such change leaves unaffected the stochastic GWB reported in this work.

[8] We refer to Guo et al. (2011) for the explanation about the computation of bulge radius and Izquierdo-Villalba et al. (2019) for improvements performed in the calculation of the bulge size after mergers.

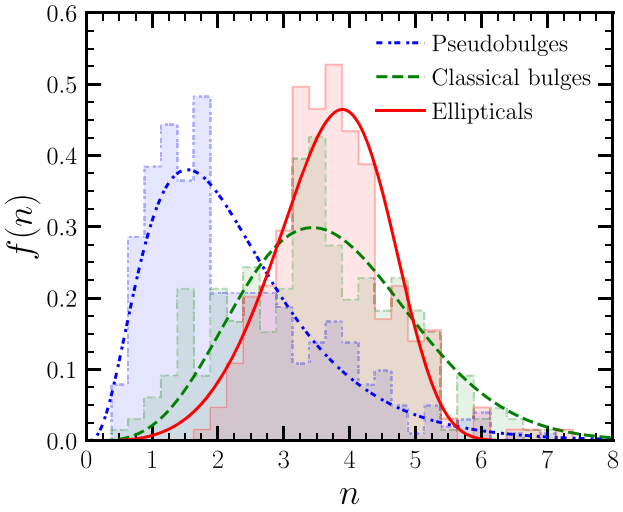

Figure 4. The histograms display the Sérsic index, n, distribution extracted from Gadotti (2009): Elliptical structures, classical bulges, and pseudo-bulges are displayed in red (solid line), green (dashed line), and blue (dot–dashed line), respectively. The solid lines display the fits to these histograms according to equation (13). The Sérsic index correlates with the central concentration of the bulge. In particular, the smaller is n, the less centrally concentrated is the bulge.

Table 1. Parameters for elliptical, classical, and pseudo-bulges from equation (13).

Bulge type	A	n_0	β
Elliptical	1.15 ± 0.04	4.24 ± 0.09	5.75 ± 0.71
Classical bulge	0.60 ± 0.13	2.47 ± 0.41	1.88 ± 0.31
Pseudo-bulge	0.021 ± 0.005	0.166 ± 0.08	0.71 ± 0.09

to L-Galaxies bulges is as follows: Each time a galaxy develops/increments the bulge via DI (minor, major merger), we extract a Sérsic index from the pseudo-bulge (classical bulge, elliptical) fit. If the galaxy had an already existing bulge, the final Sérsic index is computed as the mass-weighted average of the old bulge and the extra mass added to it. We highlight that the observations of Gadotti (2009) only take into account galaxies with stellar mass $>10^{10}\,\mathrm{M}_{\odot}$, removing from the sample dwarf galaxies. In this work, we assume that the fits presented in Table 1 hold at any stellar mass. We further assume that the $z = 0$ Sérsic indexes distribution of pseudo-bulges, classical bulges, and ellipticals hold at higher redshifts. This is a simplification and such values might evolve in the real Universe. Never the less, the results of Shibuya, Ouchi & Harikane (2015) suggest that star-forming galaxies do not display a redshift evolution in their median Sérsic index ($n \sim 1.5$).

3.3 Black hole triplets in galactic nuclei

As we discussed in the previous section, the lifetime of a binary system at the centre of a galaxy is fully determined by the hardening phase. However, in some instances, the efficiency of this process in shrinking the MBHB separation down to the GW phase can be very low (Milosavljević & Merritt 2001; Yu 2002; Merritt & Milosavljević 2005; Sesana, Haardt & Madau 2007). Indeed, if the hardening timescale is long enough, a third black hole in the pairing phase can reach the galaxy centre and interact with the MBHB system (Hoffman & Loeb 2007; Kulkarni & Loeb 2012). If this happens, the interaction between the three MBHs can lead to the prompt coalescence of two of them or a scattering event (usually ejecting the lighter MBH). Indeed, Bonetti et al. (2018a) demonstrated that these interactions are a plausible mechanism for triggering a merger in stalled binaries. In this work we treat the triple black hole interaction by including in L-Galaxies the model of Bonetti et al. (2018b). In particular, we use the Bonetti et al. (2018b) tabulated values to select those triple interactions that lead to the merger of a pair of MBHs and those causing the ejection of the lighter MBH from the system. In this latter case, the separation of the leftover MBHB is computed following Volonteri et al. (2003) and the final eccentricity is select as a random value in the range [0–1]. This grid model of Bonetti et al. (2018a) needs as an input three values: the mass of the primary black hole, the binary mass ratio, and $M_{\mathrm{BH},1}/(M_{\mathrm{BH},2} + M_{\mathrm{BH},3})$ (where $M_{\mathrm{BH},3}$ is the mass of the intruder black hole).

3.4 The growth of pairing black holes and hard binaries

The recent hydrodynamical simulations of merging galaxies with central MBHs by Capelo et al. (2015) showed that the secondary galaxy suffers large perturbations during the pericentre passages around the central one. In these circumstances, the black hole of the secondary galaxy experiences accretion enhancements, mainly correlated with the galaxy mass ratio. In this work, we include these findings assuming that right before the galaxy merger, the black hole of the secondary galaxy is able to generate or increase its gas reservoir. In this work we determine the amount of mass deposited in the MBH reservoir according to equation (1). The growth in this pairing phase is modelled in the same way as we did for nuclear black holes, i.e. the accretion rate is determined by an initial Eddington limited phase followed by a self-regulated growth in which the black hole consumes the gas at low Eddington rates (see Izquierdo-Villalba et al. 2020, for the equations that govern that growth phase).

Gas accretion on to MBHB systems has been extensively studied during the last years (D'Orazio, Haiman & MacFadyen 2013; Farris et al. 2014; Moody, Shi & Stone 2019; Muñoz, Miranda & Lai 2019; D'Orazio & Duffell 2021). Despite not being a simple process to study and model, it has been possible to draw a general picture. The circumbinary disc gas is progressively stripped from its inner edges, feeding trough accretion streams mini-disc around the two MBHs that ultimately are accreted. Interestingly, it has been shown that irrespective of the mass ratio of the binaries, the gas accretion on to the secondary black hole is sufficient to change the final mass ratio of the binary, moving the initial values toward larger ones (see e.g Farris et al. 2014; Duffell et al. 2020). Based on this picture, during the hardening phase of the MBHB system, we follow the results of Duffell et al. (2020). Accordingly, the accretion rate of a primary black hole ($\dot{M}_{\mathrm{BH}_1}$) is fully determined by the binary mass ratio (q) and the accretion rate of the secondary black hole ($\dot{M}_{\mathrm{BH}_2}$):

$$\dot{M}_{\mathrm{BH}_1} = \dot{M}_{\mathrm{BH}_2}(0.1 + 0.9q). \tag{14}$$

Therefore, each time a hard binary has formed surrounded by an circumbinary accretion disc, we fix the accretion of the secondary black hole at the Eddington limit and determine the accretion on to the primary based on equation (14).

4 RESULTS

In this section, we present the main results. We infer from L-Galaxies the chirp mass distribution of the MBHBs and merger rates from the model. We then report on the predictions for the amplitude of the GWB at the ∼n-Hz frequencies proved by the PTA

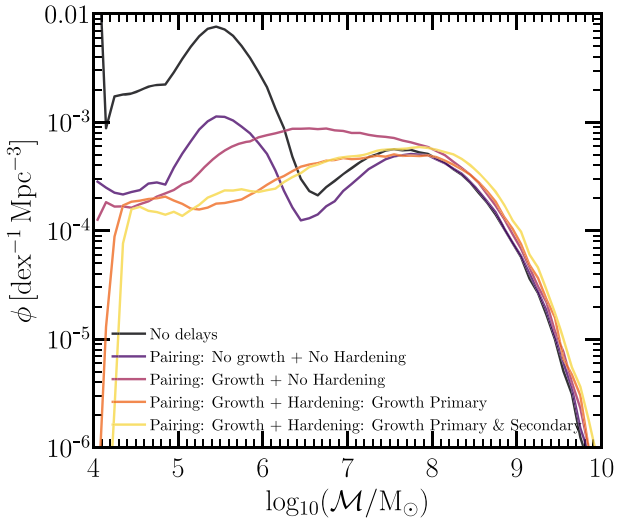

Figure 5. Chirp mass function of the *merged* black holes. The black line refers to the case when no delays are assumed in the model. Violet and purple lines represent the chirp mass function with only the delay in the pairing phase with and without gas accretion on to the in-spiralling black holes, respectively. Orange (yellow) line represents the model with both pairing and hardening phase with the secondary black hole in the binary able (unable) to accrete matter form the circumbinary disc.

experiments. Finally, we generate two variants of the model where the GWB amplitude is increased by pushing the gas accretion on to the MBHs after mergers and DIs (see equations 1 and 2). We explored the capability to produce a population of MBHs compatible with current constraints from observational works. The results on the amplitude of the GW stochastic background is tested against current knowledge on the AGN and MBH mass distributions recalling that the model of BH growth and spin evolution of `L-Galaxies` has been calibrated to be consistent with this set of observations (see Izquierdo-Villalba et al. 2020).

4.1 Merged black holes: chirp masses and merger ratios

The chirp mass of a binary, in the source frame, is the quantity that takes an important role in the amplitude of the GW emitted by a coalescing binary and is defined as

$$\mathcal{M} = \frac{(M_{\rm BH,1}M_{\rm BH,2})^{3/5}}{(M_{\rm BH,1}+M_{\rm BH,2})^{1/5}}. \quad (15)$$

In Fig. 5, we present the *rest-frame* chirp mass function of merged black holes. When no delays are included in the model, we can see a large population of mergers with $\mathcal{M} < 10^6\,\mathrm{M}_\odot$. This is an artefact produced by the seeding model, where all the newly resolved galaxies are seeded with a fix $10^4\,\mathrm{M}_\odot$ seed black hole. Nevertheless, when a dynamical friction time-delay is added in the pairing phase (without any hardening phase), the merger rate of low-mass binary systems is reduced. In particular, we can see that below $\mathcal{M} \lesssim 10^6\,\mathrm{M}_\odot$, the mass function has decreased by a >1 dex. Interestingly, the pairing phase does not have an effect on the high-mass end of the distribution, where no significant differences are found. This is caused by the fact that MBHs of $M_{\rm BH} > 10^7\,\mathrm{M}_\odot$ have a short pairing time-scale, typically <100 Myr (see Fig. 3).

In the same Fig. 5, we explored the effects of black hole mass-growth during the pairing phase. We refer the reader to Section 3.4 for the treatment used to deal with MBH growth in the pairing phase. Notice that, as we did before, no hardening is added yet. Therefore, as soon as the pairing phase is over, an MBH merger takes place. As shown in the figure, the main difference between the model with and without growth is that the former gives a larger number of events at $\mathcal{M} > 10^6\,\mathrm{M}_\odot$. This different behaviour is caused by the effectiveness of the growth during the pairing phase in reducing the mass difference between the pairing black hole and nuclear MBH at the time of the binary formation and its subsequent coalescence (Capelo et al. 2015). Interestingly, the larger differences are found at $\mathcal{M} < 10^{7.5}\,\mathrm{M}_\odot$ that arise from the fact that secondary MBHs involved in these mergers display $M_{\rm BH} < 10^6\,\mathrm{M}_\odot$ and $t_{\rm dyn}^{\rm BH} \lesssim 1$ Gyr. Such large time delays allow these MBHs to consume all (or most of) the gas reservoir stored during the pre-merger phase. In contrast, at $\mathcal{M} > 10^{7.5}\,\mathrm{M}_\odot$, the secondary MBHs ($M_{\rm BH} > 10^6\,\mathrm{M}_\odot$) display $t_{\rm dyn}^{\rm BH} \lesssim 0.1$ Gyr, having less time to increase their masses before the coalescence. When the hardening phase is included on top of the pairing one, the chirp mass function changes principally at $\mathcal{M} < 10^7\,\mathrm{M}_\odot$ where the amplitude decreases a factor of ∼4. On the other hand, the massive end is almost untouched. This different mass behaviour is just the natural consequence of the evolution of hard binaries in Sérsic model profiles. As discussed in Section 3.2, the larger is the mass of the binary system the smaller is the hardening time-scale (see Biava et al. 2019). Particularly, MBHB systems with total mass $M_{\rm bin} = 10^9\,\mathrm{M}_\odot$ display a hardening time-scale ∼0.2 Gyr, whereas for $M_{\rm bin} = 10^6\,\mathrm{M}_\odot$, the time increases up to ∼10 Gyr. Thus, the decay of the mass function at $\mathcal{M} < 10^7\,\mathrm{M}_\odot$ is the effect of the MBHBs stalling at the hardening phase.

The hardening phase explored before only allows accretion on to the primary black hole during the lifetime of the MBHB system. However, as discussed in Section 3.4, we included the possibility of the secondary MBH to accrete matter from the cirbumbinary disc that surrounds the binary system. In Fig. 5, we present the chirp mass function for that case. As shown, no big differences are seen at $\mathcal{M} < 10^8\,\mathrm{M}_\odot$ when we compare the hardening model with and without the growth of the secondary MBHs. The larger differences are displayed in the massive end ($\mathcal{M} > 10^8\,\mathrm{M}_\odot$), where there is a clear increase of the mass function for the hard model with gas accretion on to the secondary MBH. This effect has been also seen in some recent works based on the post-processing of hydrodynamics simulations. For instance, Siwek et al. (2020) found that boosting the growth of the secondary black hole over the primary one causes a shift of the chirp mass function towards large masses.

In Fig. 6, we analyse the effect of different delays and gas accretion prescriptions on the distribution of merging binaries in the ($M_{\rm BH,1}$, $M_{\rm BH,2}$) plane. In the first panel, we present the results when no delays are added. For $M_{\rm BH,1} > 10^5\,\mathrm{M}_\odot$, a large number of mergers happen with seed mass black holes, causing that the most of the primary MBHs ($M_{\rm BH,1} > 10^7\,\mathrm{M}_\odot$) display merger ratios $q < 10^{-3}$ (dark horizontal black stripe at the bottom of the panel). Despite this, the models finds a significant number of events with $M_{\rm BH,2} > M_{\rm seed}$ and $q > 0.01$, although a large scatter is seen, especially at $M_{\rm BH,1} < 10^8\,\mathrm{M}_\odot$. The second panel of the Fig. 6 presents the same but when the pairing phase is added. No big differences are see, except a large decrease of the mergers involving seed mass MBHs (see Fig. 5 to see better such drop). When we add the growth in the pairing phase (third panel of Fig. 6), we see significant changes. In particular, at $10^7 < M_{\rm BH,1} < 10^8\,\mathrm{M}_\odot$, the mergers happen with more massive secondary black holes. In this mass range, the mass of the involved secondary MBH displays a bi-modality. There is a big cloud at $10^5 < M_{\rm BH,2} < 10^6\,\mathrm{M}_\odot$, which prompt mergers with $0.001 < q < 0.1$. As we already discussed, such secondary MBHs increased their final q values at the coalescence time thanks to their

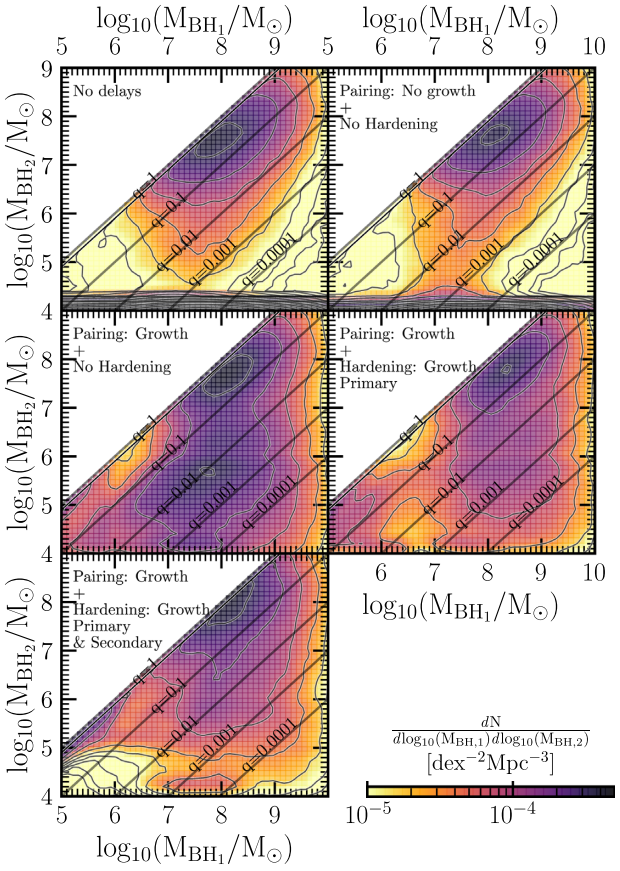

Figure 6. Relation between the mass of the primary and secondary black hole ($M_{\rm BH_1}$ and $M_{\rm BH_2}$, respectively). Different black lines highlight different binary mass ratios, $q = 1, 0.1, 10^{-2}, 10^{-3}, 10^{-4}$. Upper left-hand panel: predictions when no delays in the black hole mergers are assumed. Upper right-hand panel: delay due to the pairing phase. Middle left-hand panel: accretion of the remaining accretion disc is allowed during the pairing phase. Middle right-hand panel: delay by both pairing and hardening phase. The primary black hole consumes the whole circumbinary disc. Lower left-hand panel: delay by both pairing and hardening phase. The secondary black hole is able to accrete part of the circumbinary disc.

large pairing times ($t_{\rm dyn}^{\rm BH} \lesssim 1$ Gyr) that allow them to consume most of the gas reservoir stored during the pre-merger phase. On the other hand, we can see a secondary cloud at $10^{7.5} < M_{\rm BH,2} < 10^{8}\,{\rm M}_\odot$. Although it was already present in the pairing model without growth, in this case, the number of events has increased. Although the typical merger ratios ($q > 0.1$) are more shifted towards $q = 1$, no large differences are seen with respect to the ones of the pairing phase without growth. As commented before, the small pairing times of these secondary MBHs ($t_{\rm dyn}^{\rm BH} \lesssim 0.1$ Gyr) disfavour large mass changes during the pairing phase. When a hardening phase is added (fourth panel of Fig. 6), a large number of mergers with $10^{5} < M_{\rm BH,2} < 10^{6}\,{\rm M}_\odot$ vanishes. In this case, the merger ratios that predominate are the ones $q > 0.1$. Finally, when we allow the growth of the secondary black hole during the hardening phase (fifth panel of Fig. 6), we see an effect of systematically increasing the q parameter regardless the value of $M_{\rm BH,1}$. Indeed, in this case, most of the mergers with $M_{\rm BH} > 10^{8}\,{\rm M}_\odot$ have $q \sim 1$. As discussed before, this effect is also seen by Siwek et al. (2020), who, exploring different growth models, found that rising the mass accreted by the secondary black hole causes an increase of black hole merger events close to $q = 1$.

From here on, we will consider our *fiducial model* to be the one in which growth is allowed in both pairing and hardening. Specifically, during the hardening phase, we allow both *primary* and *secondary* MBH to accrete matter from the circumbinary disc.

4.2 The GW stochastic background

Following Sesana et al. (2008), the characteristic stochastic GWB from a population of inspiralling MBHBs can be expressed as

$$h_{\rm c}^2(f) = \frac{4G^{5/3}}{f^2c^2\pi}\int\int \frac{{\rm d}z{\rm d}\mathcal{M}}{(1+z)}\frac{{\rm d}^2n}{{\rm d}z{\rm d}\mathcal{M}}\frac{{\rm d}E_{\rm GW}(\mathcal{M})}{{\rm d}\ln f_{\rm r}}, \tag{16}$$

where ${\rm d}^2n/{\rm d}z{\rm d}\mathcal{M}$ is the comoving number density of MBHB merger per unit redshift, z, and rest-frame chirp mass, $\mathcal{M}$, and f is the frequency of the GWs in the observer frame. The quantity ${\rm d}E_{\rm GW}/{\rm dln}f_{\rm r}$ represents the differential energy spectrum of the binary, i.e. the energy emitted per logarithmic rest-frame frequency, $f_{\rm r}$. Given that we are interested in the population of inspiral MBHB in the PTA band, we make the specific assumption that the MBHBs producing the GWB are in perfect circular orbits evolving purely due to GW emission. From these assumptions, equation (16) can be re-written as

$$h_{\rm c}^2(f) = \frac{4G^{5/3}f^{-4/3}}{3c^2\pi^{1/3}}\int\int {\rm d}z{\rm d}\mathcal{M}\frac{{\rm d}^2n}{{\rm d}z{\rm d}\mathcal{M}}\frac{\mathcal{M}^{5/3}}{(1+z)^{1/3}}, \tag{17}$$

which is often **expressed** as

$$h_{\rm c}(f) = A\left(\frac{f}{f_0}\right)^{-2/3}, \tag{18}$$

where A is the amplitude of the signal at the reference frequency f_0. Usually, the GWB amplitude is referred at $f_0 = 1{\rm yr}^{-1}$. Hereafter, we will denote $A(f_0 = 1{\rm yr}^{-1})$ as $A_{\rm yr^{-1}}$. In Fig. 7, we present the model predictions. The value of $A_{\rm yr^{-1}}$ corresponds to $\sim 1.2\times10^{-15}$, being in agreement with the upper limits placed by the EPTA (Lentati et al. 2015), the NANOGrav (Arzoumanian et al. 2018), the PPTA (Shannon et al. 2015), and the IPTA (Verbiest et al. 2016) projects. The model is also compatible with the predictions coming from the bulge–black hole relation in the local Universe (Sesana et al. 2016). Other works based on SAMs or hydrodynamics simulations displayed similar results. For instance, Kelley et al. (2017a), by using the `Illustris` simulation, reported $A_{\rm yr^{-1}} = 7.1 \times 10^{-16}$. Despite the good agreement with other works, Fig. 7 shows that our predictions are below the most recent results of NANOGrav (12.5-yr data analysis; Arzoumanian et al. 2020), PPTA (DR2; Goncharov et al. 2021), and EPTA (DR2; Chen et al. 2021). Section 4.3 will be devoted to the comparison between our predictions and NANOGrav/PPTA/EPTA latest results, trying to reconcile theoretical results with observational constraints.

In the middle panel of Fig. 7, we show the GW spectrum signal produced by binary systems of three different chirp masses: $10^{7} < \mathcal{M} < 10^{8}\,{\rm M}_\odot$, $10^{8} < \mathcal{M} < 10^{9}\,{\rm M}_\odot$, and $\mathcal{M} > 10^{9}\,{\rm M}_\odot$. As shown, the two latter bins contribute the most to the signal. On the other extreme, binaries of $10^{7} < \mathcal{M} < 10^{8}\,{\rm M}_\odot$ have a marginal effect, contributing typically 1 dex less than $\mathcal{M} > 10^{8}\,{\rm M}_\odot$. Regarding the mass ratios of MBHBs generating the GWB, the bottom panel of Fig. 7 shows that systems with $q > 0.1$ are the ones producing most of the signal. Furthermore, the results show that the smaller the q parameter, the smaller is the effect of the binary system in the GWB. For instance, binary systems with $0.01 < q < 0.1$ and $q < 0.01$ generate, respectively, 0.4 and 0.13 times smaller amplitude

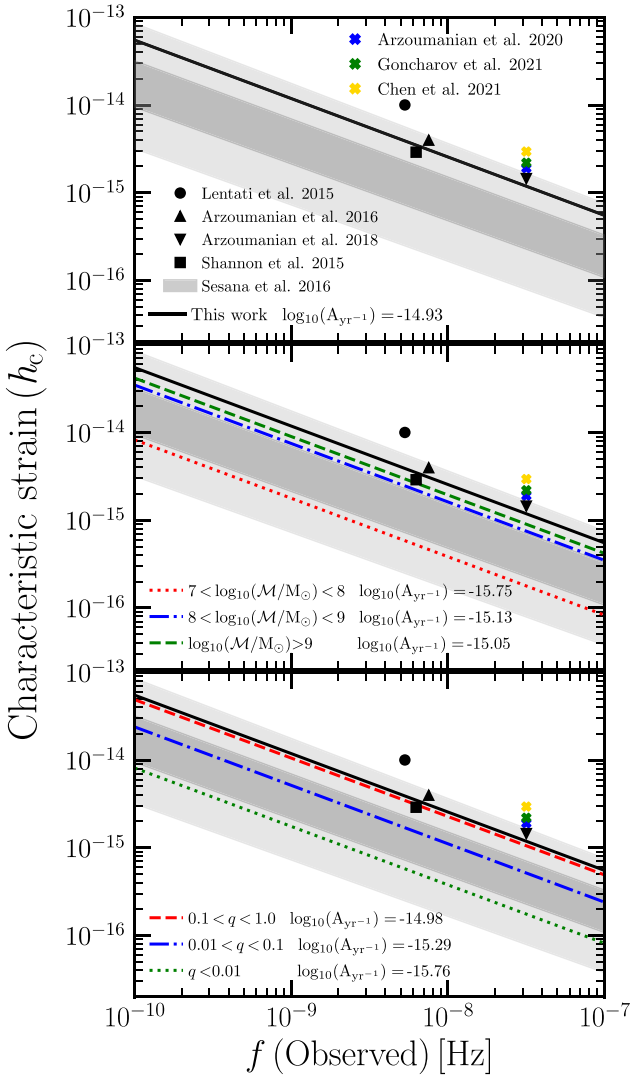

Figure 7. Upper panel: GWB amplitude predicted by the model. Middle panel: GWB signal computed in three different chirp mass intervals: $10^7 < \mathcal{M} < 10^8\,M_\odot$ (red dotted line), $10^8 < \mathcal{M} < 10^9\,M_\odot$ (blue dot–dashed line), and $\mathcal{M} > 10^9\,M_\odot$ (green dashed line). Lower panel: GWB signal same as above split in three binary mass ration: $0.1 < q < 1.0$ (red dashed line), $0.01 < q < 0.01$ (blue dot–dashed line), and $q < 0.01$ (green dotted line). In all the three plots, the circle, triangle, and square points are the upper limits placed by the EPTA (Lentati et al. 2015), the NANOGrav (Arzoumanian et al. 2016, 2018), and the PPTA (Shannon et al. 2015) projects, respectively. Blue, green, and yellow crosses at $f(\text{Observed}) = 1\text{yr}^{-1}$ represents, respectively, the measurements of the common red noise reported by Arzoumanian et al. (2020), Goncharov et al. (2021), and Chen et al. (2021). The shaded areas show the constrains coming from the local Universe bulge–black hole relation (Sesana et al. 2016): Dark and clear grey areas represent the 1σ and 2σ confidence interval.

than the total signal. These results are consistent with Sesana et al. (2008) and Sesana (2013), who showed that >95 per cent of the GW signal at ∼ n-Hz frequencies comes from BH major mergers ($q > 0.25$) involving BHs with mass $>10^8\,M_\odot$ at $z < 1.5$. Similar results were recently reported by Casey-Clyde et al. (2021). By using empirical relations for quasar luminosity functions, quasar lifetime, and MBHB mass ratio distribution, the authors concluded that most of the GWB signal would be produced by MBHBs of mass $>10^8\,M_\odot$ at $z \sim 0.5$.

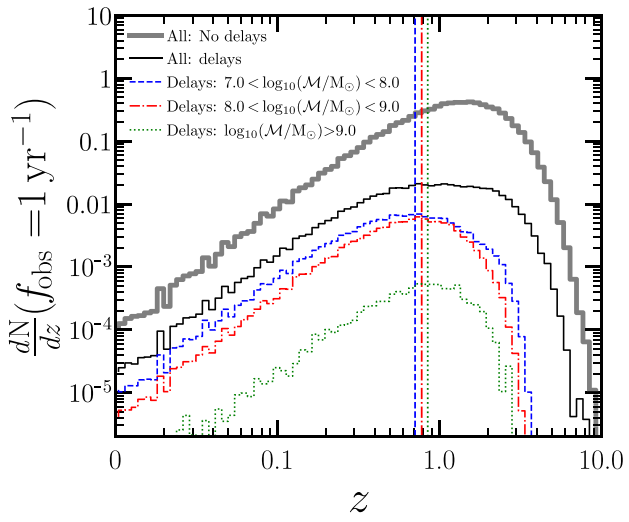

Figure 8. Merger rates predicted by `L-Galaxies`. The black thick line displays the predictions of the model when no delays (pairing and hardening) are assumed. Think black thick line represents the same but with pairing and hardening delays. Coloured lines represent the merger ratios in the model with delays at different chirp masses: $7 < \log_{10}(\mathcal{M}/M_\odot) < 8$ (blue dashed line), $8 < \log_{10}(\mathcal{M}/M_\odot) < 9$ (red dot–dashed line) and $\log_{10}(\mathcal{M}/M_\odot) > 9$ (green dotted line). The vertical lines highlight the maximum of each distribution.

In Fig. 8, we present the merger rates for MBHB without any binary treatment (thin black line) and when we included the pairing and hardening delay (think black line). The figure shows that the MBHB model causes a large change in the rates at which the MBHs coalesce. Whereas the integrated merger rate without MBH merger delays reaches up to $1.03\,\text{yr}^{-1}$, in the version with delays, it drops down to $0.06\,\text{yr}^{-1}$. For the latter case, we have explored the predictions for $\mathcal{M} > 10^7\,M_\odot$. As we can see, the mergers of these massive binaries happen at relatively low-z, being typically at $z \sim 1$. When the population is divided into different mass bins, a mild redshift difference is seen, with larger $\mathcal{M}$ being the systems that merge slightly earlier. Besides, at $z < 1$, merger events of $\mathcal{M} > 10^9\,M_\odot$ binaries decrease faster than the ones of $10^7 < \mathcal{M} < 10^8\,M_\odot$ and $10^8 < \mathcal{M} < 10^9\,M_\odot$, which have similar behaviour.

4.3 The stochastic gravitational background confronting the mass and quasar luminosity functions

Recently, by using the 12.5-yr pulsar-timing data set of NANOGrav collaboration Arzoumanian et al. (2020) reported strong evidences of a stochastic process with $A_{\text{yr}^{-1}}$ spanning between 1.37×10^{-15} and 2.67×10^{-15} (5–95 per cent quantiles) and median value of 1.92×10^{-15}. Similar signal was also recently reported by the PPTA ($A_{\text{yr}^{-1}} \sim 1.9 \times 10^{-15}$–$2.6 \times 10^{-15}$) and EPTA ($A_{\text{yr}^{-1}} \sim 2.23 \times 10^{-15}$–$3.8 \times 10^{-15}$) second data release (Chen et al. 2021; Goncharov et al. 2021). Even though such signals did not display significant evidences of quadrupolar correlations needed to claim GW detection, it is interesting to test which are the predictions of our model for such large GW signal. Specifically, in this section, we present the model predictions when reaching a GWB compatible with the median value and 95 per cent quantiles of Arzoumanian et al. (2020). The conclusions presented in this section are the same when the limits of Goncharov et al. (2021, PPTA) and Chen et al. (2021, EPTA) are used.

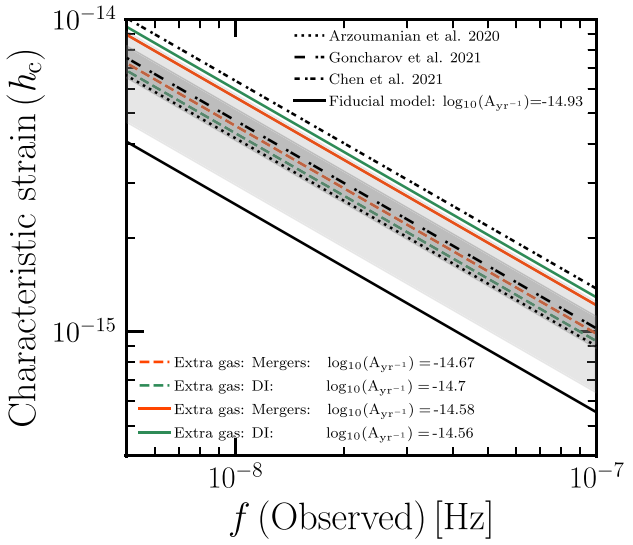

Figure 9. Amplitude of the GWB in the frequency range 10^{-10}–10^{-7} Hz. While the solid black line represents the fiducial model, orange and green lines display the results when the gas accretion during merger and DIs is boosted, respectively. Dashed (solid) lines represent the model predictions when $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ ($A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$) is reached. The clear grey shaded area and the dotted line show the constrains coming from Arzoumanian et al. (2020). The dark grey shaded area and the long dash–dotted line show the constrains coming from Goncharov et al. (2021). The short dash–dotted line represents the results of Chen et al. (2021) (to avoid confusion, we did not show the upper and lower limits of Chen et al. 2021).

To increase the GW signal, we explored two variants of the model. The first one consisted in increasing the amount of gas accreted by the black holes during galaxy mergers (hereafter model *increased merger*, IM) by increasing the parameter $f_{\rm BH}^{\rm merger}$ (see equation 1) by a factor of 2 and 3 to reach $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ and $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$, respectively. In the second variant of the model, we left the mergers untouched and changed the gas accretion during DIs (hereafter model *increased DI*, IDI). Specifically, we increase $f_{\rm BH}^{\rm DI}$ (see equation 2) by a factor of 9 and 20 to achieve, respectively, a GWB of $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ and $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$. The GWBs produced by these four model variants are presented in Fig. 9.

The question to answer now is whether these new models are also consistent with constraints on the black hole mass and luminosity function. In Fig. 10, we present the comparison between the models and the current observations of the black hole mass function (BHMF) in the local Universe (Marconi et al. 2004; Shankar et al. 2004, 2009, 2013). As shown, our fiducial run is in good agreement with these observations. On the other hand, the models with a boosted mass growth display values in tension with the observations, especially the ones with GWB of $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$. Regardless of the GWB level, in the IDI cases, we see a behaviour compatible with observations for $10^6 < M_{\rm BH} < 10^8\,{\rm M}_\odot$. However, the massive end ($M_{\rm BH} > 10^8\,{\rm M}_\odot$) is typically overpredicted by almost a factor of 3 for $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ and ∼1 dex for $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$. A similar trend is observed in the IM models. Additionally, the latter show incompatibilities at lower masses as well ($M_{\rm BH} \sim 10^{6.5}\,{\rm M}_\odot$). Even though the high-mass tail of the BHFM seems to be not fully constrained by observations and there is still room for further improvements, some authors pointed out that current MBH mass estimates might be biased high. For instance, as reported by Shankar et al. (2016), the discrepancy between Shankar et al. (2013) and Marconi et al. (2004) might be caused by biases affecting the observations. Shankar et al. (2016) argue that, because of selection effects, the normalization of the scaling relations used to relate the black hole mass with galaxy properties (such as bulge mass and velocity dispersion) might be increased by a factor as high as $\gtrsim 3$ (see also Bernardi et al. 2007; Shankar et al. 2019). Therefore, this will yield a lower amplitude in the empirical relations that would cause smaller measurements of BH masses and BH mass density, consistent with the current non-detection of this signal by pulsar timing array experiments (see Sesana et al. 2016).

In the lower panels of Fig. 10, we show the mass function of active MBHs, selected as those with Eddington ratios larger than 0.01. The predictions are compared with Greene & Ho (2007) and Schulze & Wisotzki (2010), which performed the same Eddington ratio selection. As shown, regardless of the GWB amplitude, the fiducial, IM, and IDI models are consistent with the predictions at $10^7 < M_{\rm BH} < 10^8\,{\rm M}_\odot$. However, the IM models overpredict the population of active MBHs at $M_{\rm BH} > 10^{8.5}$. For masses $< 10^7\,{\rm M}_\odot$, we cannot draw strong conclusions when comparing predictions with observations, considering current selection effects of the latter. For instance, the flux limit imposed by Schulze & Wisotzki (2010) causes large incompleteness effects at at low black hole masses and low Eddington ratios.

In Fig. 11, we present the evolution of the quasar bolometric luminosity function (LF) from $z \sim 3$ down $z \sim 0$. Even though these functions give the number density of accreting black holes in different luminosity bins, they have been a powerful tool to extract information on how MBHs grow with cosmic time, on the geometry of the accretion discs and other fundamental quantities such as the black hole spins and radiative efficiencies. In this work, we only focus on the very bright objects, i.e. $>10^{45}$ erg s^{-1}, avoiding the comparison with lower luminosity given the current limitations on observational and theoretical models. In particular, from an observational standpoint, the covered area and depth of current surveys pose serious challenges when extracting statistical properties of the LF at the faint end (Siana et al. 2008; Masters et al. 2012; McGreer et al. 2013; Niida et al. 2016; Akiyama et al. 2018). Even more, dust attenuation effects might play an important role in shaping current measurements. On the other hand, current theoretical works show a large excesses at luminosity $<10^{45}$ erg s^{-1}. In order to reconcile observations with predictions, these works have played with empirical relations for obscuring accreting black holes or with the efficiency of the seeding process (see e.g Degraf, Di Matteo & Springel 2010; Fanidakis et al. 2012; DeGraf & Sijacki 2020). Even though these works provide interesting results shedding light on the nature of low-luminous quasars, the treatment of seeding or dust obscuration is beyond the scope of this paper. As shown in Fig. 11, the fiducial model is compatible with current observations of the quasar LF, showing a sharp cut-off at larger luminosity (Shen et al. 2020). On the other hand, the models with higher gas accretion display a completely different behaviour. Boosting the gas accretion during DI leads to a larger excess of bright quasars at $z > 1.0$. For instance, at $z \sim 2$ and for luminosities $>10^{46}$ erg s^{-1}, the models with $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ and $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$ are systematically overpredicting the number density by a factor of ∼1 and ∼2 dex, respectively. A similar behaviour is seen at $z \sim 3$. At lower redshifts ($z < 1.0$), the model follows both the fiducial results and the observed trends. This is principally caused by the decrease of important DIs events at these redshifts. Regarding the IM models, we can see similar trends at $z > 2$, where the bright end of the LF is systematically larger than the observed one. We highlight that the difference is larger with $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$. Interestingly, the excess with respect to the

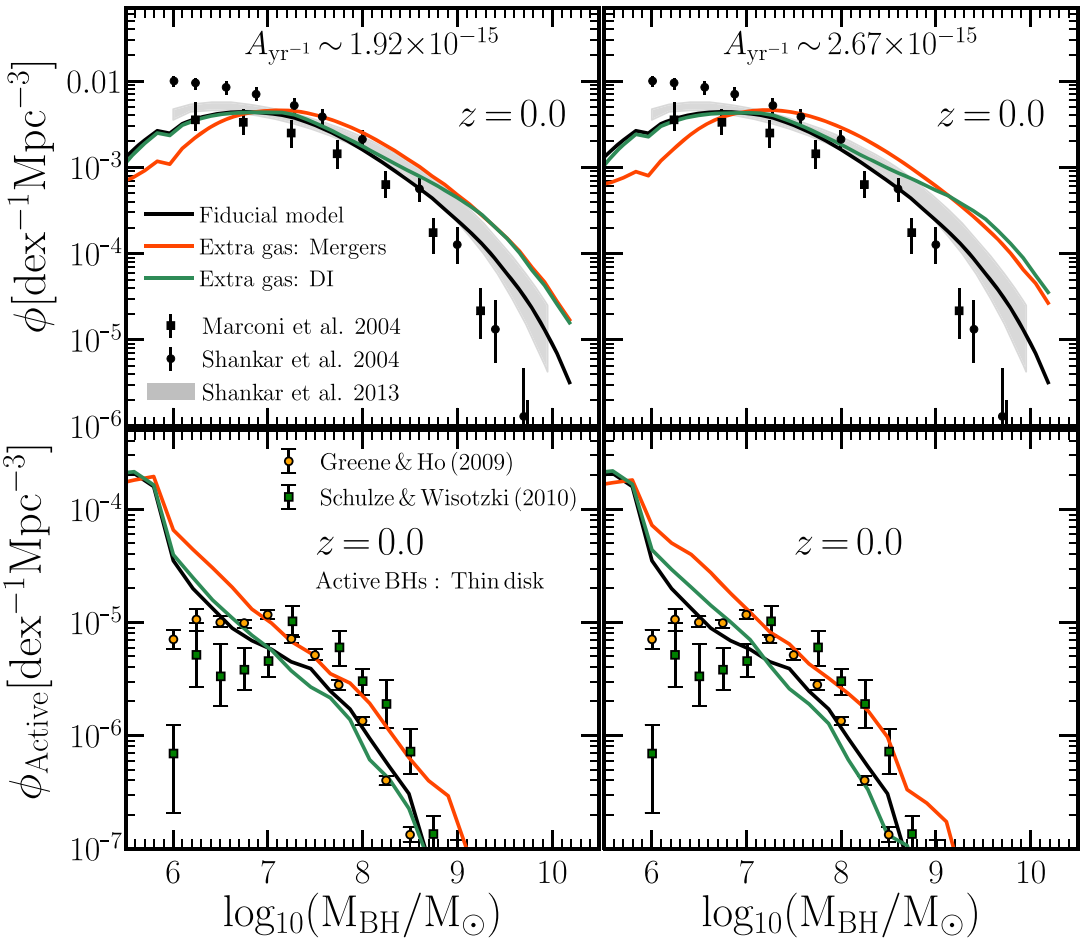

Figure 10. Model predictions when a stochastic GWB of $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ (left-hand panels) and $A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$ (right-hand panels) is reached. The upper panel display the $z = 0$ black hole mass function compared to the observational results of Marconi et al. (2004), Shankar et al. (2004), and Shankar, Weinberg & Miralda-Escudé (2013). The lower panels show the black hole mass function at $z \sim 0$ for active black holes (Eddington ratio $> 10^{-2}$) from Greene & Ho (2007) and Schulze & Wisotzki (2010) are added for comparison. In all the plots, the black line corresponds to the predictions of the fiducial model. Orange and green lines represent the results when we boost the gas accretion during mergers and DIs, respectively.

observations is smaller than with the IDI model. This is principally caused by the fact that DI events are more important than mergers at these redshifts (Izquierdo-Villalba et al. 2020). At lower redshifts, we can see larger differences with respect to the fiducial and the IDI models: IM model is systematically overprotecting the bright end of the LF ($>10^{46}$ erg s^{-1}). Such differences can be a factor of 3 (1.5) by $z \sim 0$ up to a factor of 5 (2) at $z \sim 0.5$ for $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ ($A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$).

Based on the results presented in Figs 10 and 11, we can draw the conclusion that large GWBs can be reached by our SAM just by changing the gas accretion of the black holes after mergers or DIs. However, these amplitudes are difficult to reconcile with observational constrains such as the black hole mass function or quasar bolometric luminosity function. Therefore, we highlight that the reliability of GWBs produced by both SAMs or hydrodynamical simulations must be tested by checking the properties of the full black hole population such as luminosity functions or mass distribution across cosmic time. On this line, we can find the recent work of Casey-Clyde et al. (2021) in which it is presented a new model to constrain the population of MBHB based on GWBs and quasar populations. According to the number density of quasars and their expected lifetime (Hopkins et al. 2006a, 2007), the authors pointed out that the last NANOGrav GW signal would suggest a local number density of MBHB 5 times larger than the previously detected, being 25 per cent of the MBHB system associated with quasars.

5 SUMMARY AND CONCLUSIONS

In this paper, we presented a model tracking the formation and evolution of MBHBs across cosmic time. We made use of the `L-Galaxies` SAM (Henriques et al. 2015) run on the `Millennium` DM merger trees whose mass resolution allows to draw solid conclusions for galaxies of mass $\gtrsim 10^{8.5}$ M$_\odot$ and MBHs $\gtrsim 10^{6}$ M$_\odot$. The MBHB model was developed as an extension of the work presented in Izquierdo-Villalba et al. (2020), where detailed prescriptions for the mass growth and spin evolution of MBHs were included in `L-Galaxies`. In a nutshell, the MBHs are allowed to grow trough cold gas accretion, hot gas accretion and mergers with other black holes. Specifically, the former channel is the main driver of the black hole growth and it is triggered by both galaxy mergers and DIs. During any growth events, the code tracks the evolution of the black hole spin in a self-consistent way.

Following the standard scenario, we included three different stages for the dynamical evolution of MBHBs that needs to be tracked

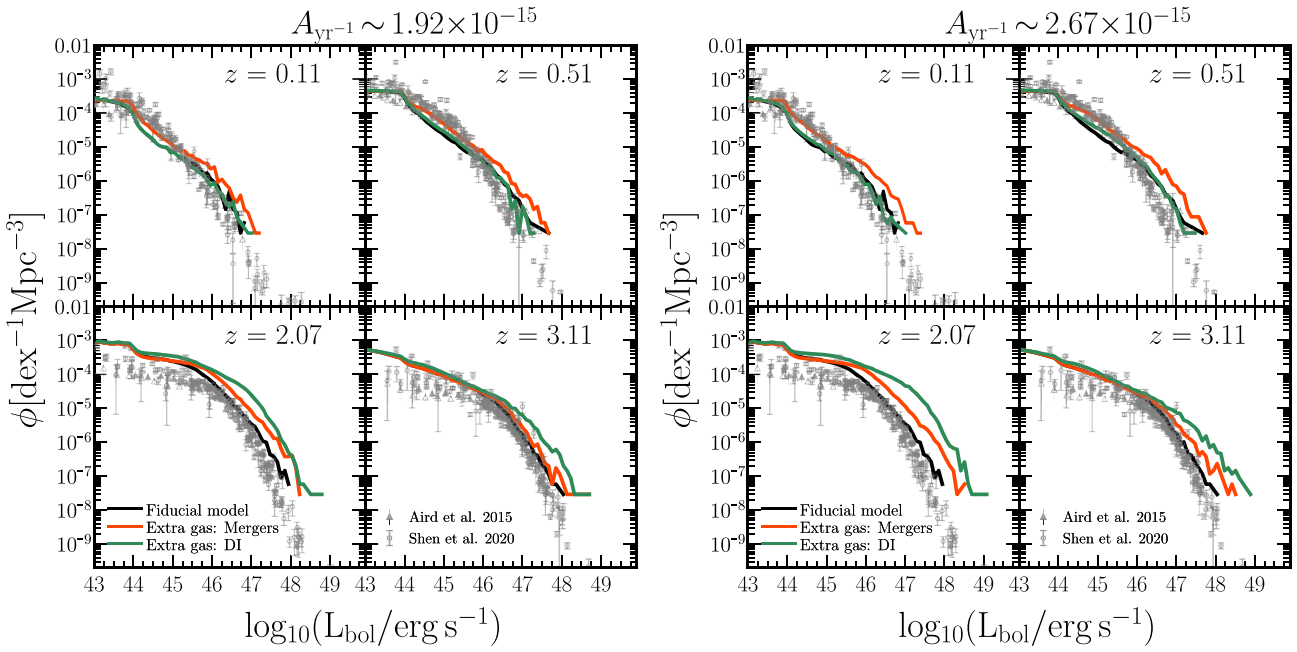

Figure 11. Quasar bolometric luminosity functions ($L_{\rm bol}$) at $z \approx 0.1$, 0.5, 2.0, 3.0. Luminosity functions are compared with the data of Shen et al. (2020) (circles) and Aird et al. (2015) (triangles). The left-hand (right-hand) panels correspond to the model predictions when a stochastic GWB of $A_{\rm yr^{-1}} \sim 1.92 \times 10^{-15}$ ($A_{\rm yr^{-1}} \sim 2.67 \times 10^{-15}$) is reached. In all the plots, the black line corresponds to the predictions of the fiducial model. Orange and green lines represent the results when we rise the gas accretion during mergers and DIs, respectively.

in order to build a population of MBHBs: *pairing*, *hardening*, and GW phase. We assumed that the first phase starts after the galaxy–galaxy merger is completed, and corresponds to the sinking process of the MBH of the satellite galaxy towards the centre of the newly formed galaxy. The process is driven by dynamical friction acting on the black holes individually, and exerted by the galaxy's stellar component. The time spent by the MBH of the less massive galaxy during the pairing phase has been computed following recent refinements of the Chandrasekhar (1943) formula, which account for the eccentricity of the MBH orbit. Since the dynamical friction time-scale depends on the initial position of the MBH relative to the host galaxy, this distance has been computed accounting for mass stripping of the secondary by the tidal field of the primary galaxy. The model has shown that the orbit of a large fraction of MBHs stalls in this phase, being a bottleneck for the formation of a bound MBHB system. Despite that, the number of MBHs with $>10^6\,\rm M_\odot$ reaching the galaxy nucleus increases towards low-z. On top of this, we have found that elliptical galaxies at $z < 1.0$ are the preferred birthplaces of MBHB systems.

During the pairing phase, we allowed the black holes to accrete their pre-merger gas reservoir. Interestingly, this has an imprint on the final chirp mass function (CBHMF) of merged MBHs. The main effects are seen at $10^6 < \mathcal{M} < 10^{7.5}\,\rm M_\odot$, where the CBHMF amplitude increases with respect to the case in which gas accretion is suppressed. Such change is due to the long-lived phase of dynamical friction ($\lesssim$1 Gyr) experienced by the MBHs in these mergers. This has led to a significant increase of their masses, by consuming all the gas reservoir stored during the pre-merger phase. A similar trend is seen at $\mathcal{M} > 10^{7.5}\,\rm M_\odot$, but the effects are smaller, given the shorter time-scales involved in these cases, which disfavoured large mass increases during the pairing phase.

When the pairing phase has ended, the MBHs form a binary system governed by the hardening and GW phase. We distinguished between two different environments in this phase: *gas-rich* and *gas-poor*. In the former case, a circumbinary gas disc around the MBHB forms and dominates the system. The torques exerted by the disc cause the shrinking of the orbit and coalescence of the two MBHs. In this environment, the binary separation is tracked by integrating numerically the differential equation of Dotti et al. (2015). In contrast, in gas-poor environments, the hardening phase is caused by the effect of stars intersecting the MBHB orbit. These interactions are able to extract a significant amount of the MBHB energy and angular momentum through the slingshot mechanism. The binary separation and eccentricity in this type of environment are tracked by integrating numerically the differential equation of Sesana & Khan (2015), assuming a Sérsic model profile for the host galaxy (Sersic 1968). Following the findings of Duffell et al. (2020), we assumed that gas accretion during the hardening phase is determined by the binary mass ratio and the accretion rate of the secondary black hole, set to the Eddington limit. Finally, regardless of the environment, we included the Bonetti et al. (2018a) model for triplet reaction among a binary and an incoming black hole as an additional mechanism capable of driving stalled binaries to coalescence. The results show that binary hardening in gas-poor environments reduces significantly the number of MBHB merges at $\mathcal{M} < 10^7\,\rm M_\odot$ while leaving the high-mass end of the CBHMF untouched. This different mass behaviour is caused by the evolution of hard binaries in Sérsic model profiles, where the lighter MBHBs have hardening times $\sim$2 dex larger than the most massive ones.

Thanks to the large volume and mass resolution of the `Millennium` simulation, we explored the model to predict the amplitude of the stochastic GWB at the frequencies proved by the *Pulsar timing array* (PTA) experiments. The model shows an amplitude at $1\,\rm yr^{-1}$ of $A_{\rm yr^{-1}} \sim 1.2 \times 10^{-15}$, being principally produced by binary systems with $\mathcal{M} > 10^8\,\rm M_\odot$ and $q > 0.1$. The GWB reported in this work is in agreement with current upper limits provided by Lentati

et al. (2015), Arzoumanian et al. (2016), and Shannon et al. (2015), but in tension with the last constraints reported by Arzoumanian et al. (2020). Therefore, we considered the amplitude identified by Arzoumanian et al. (2020) (under the hypothesis that is a GWB coming form MBHBs) and asked what modifications to the model could produce a GWB level consistent with Arzoumanian et al. (2020) results. Only by boosting the MBH gas accretion during mergers and DIs we produced a larger GWB amplitude ($A_{\rm yr^{-1}} = 1.37 \times 10^{-15}$–$2.67 \times 10^{-15}$) more consistent with the amplitude recently reported by the NANOGrav collaboration (Arzoumanian et al. 2020). Unlike previous studies in the literature, we confronted the predictions on the amplitude of the stochastic GWB with constraints from key observations such as the quasar luminosity functions (LFs) and local black hole mass function (BHMF). In particular, large GW amplitude values ($A_{\rm yr^{-1}} > 1.92 \times 10^{-15}$) made difficult to reconcile model predictions with the observational constraints. In particular, we showed that the models with large GWB display a large excess of bright quasars at any redshift. For instance, at $z \sim 2$, quasars with luminosity $>10^{46}$ erg s^{-1} are systematically overpredicted by a factor of 2 dex. At $z < 0.5$, such overprediction is still present, especially in the model where gas accretion on to mergers was boosted. Regarding the BHMF, the models with GWBs compatible with Arzoumanian et al. (2020) constraints display values in tension with the observations, especially in the massive end ($M_{\rm BH} > 10^8$ M$_\odot$) where the difference with current observational constraints reach up to 1–1.5 dex.

The model presented here is a step forward for the study of MBHBs across cosmic time. In future, thanks to the flexibility of the model, we will extend the analysis to the `MillenniumII` DM merger trees (Springel 2005; Boylan-Kolchin et al. 2009). Their different box sizes and DM mass resolutions will offer the capability to explore the physical processes ruling the evolution of MBHs over a wider range of masses and environments. Therefore, we will be able to characterize not only the formation, evolution, and environments of the most massive binary systems accessible through PTA experiments (Kramer & Champion 2013; McLaughlin 2013; Manchester et al. 2013) but also the less massive ones proved by the *Laser Interferometer Space Antenna* (*LISA*; Amaro-Seoane et al. 2017).

ACKNOWLEDGEMENTS

DIV and AS acknowledge financial support provided under the European Union's H2020 ERC Consolidator Grant 'Binary Massive Black Hole Astrophysics' (B Massive, Grant Agreement: 818691). DIV acknowledges also financial support from INFN H45J18000450006. MC acknowledges funding from MIUR under the Grant No. PRIN 2017-MB8AEZ. SB acknowledges partial support from the project PGC2018-097585-B-C22. This work used the 2015 public version of the Munich model of galaxy formation and evolution: `L-Galaxies`. The source code and a full description of the model are available at http://galformod.mpa-garching.mpg.de/public/LGalaxies/. Finally, we thank the anonymous referee for the many suggestions that improved the quality of this paper.

DATA AVAILABILITY

The simulated data underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003, ApJ, 597, 21
Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015, MNRAS, 451, 1892
Akiyama M. et al., 2018, PASJ, 70, S34
Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)
Angulo R. E., White S. D. M., 2010, MNRAS, 405, 143
Antonini F., Barausse E., Silk J., 2015, ApJ, 812, 72
Armitage P. J., Natarajan P., 2002, ApJ, 567, L9
Arzoumanian Z. et al., 2015, ApJ, 810, 150
Arzoumanian Z. et al., 2016, ApJ, 821, 13
Arzoumanian Z. et al., 2018, ApJ, 859, 47
Arzoumanian Z. et al., 2020, ApJ, 905, L34
Bailes M. et al., 2016, MeerKAT Science: On the Pathway to the SKA
Barausse E., 2012, MNRAS, 423, 2533
Barausse E., Rezzolla L., 2009, ApJ, 704, L40
Begelman M. C., Blandford R. D., Rees M. J., 1980, Nature, 287, 307
Bernardi M., Sheth R. K., Tundo E., Hyde J. B., 2007, ApJ, 660, 267
Bertone S., De Lucia G., Thomas P. A., 2007, MNRAS, 379, 1143
Biava N., Colpi M., Capelo P. R., Bonetti M., Volonteri M., Tamfal T., Mayer L., Sesana A., 2019, MNRAS, 487, 4985
Binney J., Tremaine S., 2008, Galactic Dynamics, 2nd edn
Bonetti M., Sesana A., Barausse E., Haardt F., 2018a, MNRAS, 477, 2599
Bonetti M., Haardt F., Sesana A., Barausse E., 2018b, MNRAS, 477, 3910
Bonetti M., Sesana A., Haardt F., Barausse E., Colpi M., 2019, MNRAS, 486, 4044
Bonoli S., Marulli F., Springel V., White S. D. M., Branchini E., Moscardini L., 2009, MNRAS, 396, 423
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Capelo P. R., Volonteri M., Dotti M., Bellovary J. M., Mayer L., Governato F., 2015, MNRAS, 447, 2123
Casey-Clyde J. A., Mingarelli C. M. F., Greene J. E., Pardo K., Nañez M., Goulding A. D., 2021, preprint (arXiv:2107.11390)
Chandrasekhar S., 1943, ApJ, 97, 255
Chen S. et al., 2021, MNRAS, 508, 4970
Colpi M., 2014, Space Sci. Rev., 183, 189
Colpi M., Sesana A., 2017, in G. Auger, E. Plagnol, Gravitational Wave Sources in the Era of Multi-Band Gravitational Wave Astronomy. eds, World Scientific
Colpi M., Mayer L., Governato F., 1999, ApJ, 525, 720
Cuadra J., Armitage P. J., Alexander R. D., Begelman M. C., 2009, MNRAS, 393, 1423
D'Orazio D. J., Duffell P. C., 2021, ApJ, 914, L21
D'Orazio D. J., Haiman Z., MacFadyen A., 2013, MNRAS, 436, 2997
De Rosa A. et al., 2019, New Astron. Rev., 86, 101525
DeGraf C., Sijacki D., 2020, MNRAS, 491, 4973
Degraf C., Di Matteo T., Springel V., 2010, MNRAS, 402, 1927
Desvignes G. et al., 2016, MNRAS, 458, 3341
Dotti M., Colpi M., Haardt F., Mayer L., 2007, MNRAS, 379, 956
Dotti M., Colpi M., Pallini S., Perego A., Volonteri M., 2013, ApJ, 762, 68
Dotti M., Merloni A., Montuori C., 2015, MNRAS, 448, 3603
Dressler A., Richstone D. O., 1988, ApJ, 324, 701
Drory N., Alvarez M., 2008, ApJ, 680, 41
Drory N., Fisher D. B., 2007, ApJ, 664, 640
Duffell P. C., D'Orazio D., Derdzinski A., Haiman Z., MacFadyen A., Rosen A. L., Zrake J., 2020, ApJ, 901, 25
Dutton A. A., Macciò A. V., 2014, MNRAS, 441, 3359
Dvorkin I., Barausse E., 2017, MNRAS, 470, 4547
Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069
Elmegreen B. G., Bournaud F., Elmegreen D. M., 2008, ApJ, 688, 67
Enoki M., Inoue K. T., Nagashima M., Sugiyama N., 2004, ApJ, 615, 19
Escala A., Larson R. B., Coppi P. S., Mardones D., 2004, ApJ, 607, 765
Escala A., Larson R. B., Coppi P. S., Mardones D., 2005, ApJ, 630, 152
Faber S. M., 1999, Adv. Space Res., 23, 925
Fanidakis N. et al., 2012, MNRAS, 419, 2797
Farris B. D., Duffell P., MacFadyen A. I., Haiman Z., 2014, ApJ, 783, 134
Foster R. S., Backer D. C., 1990, ApJ, 361, 300
Gadotti D. A., 2009, MNRAS, 393, 1531
Genzel R., Townes C. H., 1987, ARA&A, 25, 377
Genzel R., Hollenbach D., Townes C. H., 1994, Rep. Prog. Phys., 57, 417
Goncharov B. et al., 2021, ApJ, 917, L19

Greene J. E., Ho L. C., 2007, ApJ, 667, 131
Guo Q. et al., 2011, MNRAS, 413, 101
Haehnelt M. G., Rees M. J., 1993, MNRAS, 263, 168
Häring N., Rix H.-W., 2004, ApJ, 604, L89
Henriques B. M. B., White S. D. M., Thomas P. A., Angulo R., Guo Q., Lemson G., Springel V., Overzier R., 2015, MNRAS, 451, 2663
Hills J. G., 1983, AJ, 88, 1269
Hobbs G. et al., 2010, Class. Quantum Gravity, 27, 084013
Hoffman L., Loeb A., 2007, MNRAS, 377, 957
Hopkins P. F., Hernquist L., Martini P., Cox T. J., Robertson B., Di Matteo T., Springel V., 2005, ApJ, 625, L71
Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006a, ApJS, 163, 1
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Di Matteo T., Springel V., 2006b, ApJ, 639, 700
Hopkins P. F., Richards G. T., Hernquist L., 2007, ApJ, 654, 731
Izquierdo-Villalba D., Bonoli S., Spinoso D., Rosas-Guevara Y., Henriques B. M. B., Hernández-Monteagudo C., 2019, MNRAS, 488, 609
Izquierdo-Villalba D., Bonoli S., Dotti M., Sesana A., Rosas-Guevara Y., Spinoso D., 2020, MNRAS, 495, 4681
Jaffe A. H., Backer D. C., 2003a, ApJ, 583, 616
Jaffe A. H., Backer D. C., 2003b, ApJ, 583, 616
Kauffmann G., Colberg J. M., Diaferio A., White S. D. M., 1999, MNRAS, 307, 529
Kelley L. Z., Blecha L., Hernquist L., 2017a, MNRAS, 464, 3131
Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2017b, MNRAS, 471, 4508
King I., 1962, AJ, 67, 471
Kocsis B., Haiman Z., Loeb A., 2012, MNRAS, 427, 2680
Kormendy J., 1983, ApJ, 275, 529
Kormendy J., 1988, ApJ, 325, 128
Kormendy J., Bender R., 1996, ApJ, 464, L119
Kormendy J., Ho L. C., 2013, ARA&A, 51, 511
Kormendy J., Richstone D., 1992, ApJ, 393, 559
Kramer M., Champion D. J., 2013, Class. Quantum Gravity, 30, 224009
Kulkarni G., Loeb A., 2012, MNRAS, 422, 1306
Lacey C., Cole S., 1993, MNRAS, 262, 627
Lee K. J., 2016, in Qain L., Li D., eds, ASP Conf. Ser. Vol. 502, Frontiers in Radio Astronomy and FAST Early Sciences Symposium 2015. Astron. Soc. Pac., San Francisco, p. 19
Lentati L. et al., 2015, MNRAS, 453, 2576
McGreer I. D. et al., 2013, ApJ, 768, 105
McLaughlin M. A., 2013, Class. Quantum Gravity, 30, 224008
Manchester R. N. et al., 2013, Publ. Astron. Soc. Aust., 30, e017
Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169
Marulli F., Crociani D., Volonteri M., Branchini E., Moscardini L., 2006, MNRAS, 368, 1269
Masters D. et al., 2012, ApJ, 755, 169
Merloni A., Heinz S., 2008, MNRAS, 388, 1011
Merritt D., Milosavljević M., 2005, Living Rev. Relativ., 8, 8
Mikkola S., Valtonen M. J., 1992, MNRAS, 259, 115
Milosavljević M., Merritt D., 2001, ApJ, 563, 34
Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution
Moody M. S. L., Shi J.-M., Stone J. M., 2019, ApJ, 875, 66
Muñoz D. J., Miranda R., Lai D., 2019, ApJ, 871, 84
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Nelson D. et al., 2018, MNRAS, 475, 624
Niida M., Nagao T., Ikeda H., Matsuoka K., Kobayashi M. A. R., Toba Y., Taniguchi Y., 2016, ApJ, 832, 208
O'Dowd M., Urry C. M., Scarpa R., 2002, ApJ, 580, 96
Perera B. B. P. et al., 2019, MNRAS, 490, 4666
Peters P. C., Mathews J., 1963, Phys. Rev., 131, 435
Peterson B. M. et al., 2004, ApJ, 613, 682
Pillepich A. et al., 2018, MNRAS, 475, 648
Planck Collaboration et al., 2014, A&A, 571, A16
Press W. H., Schechter P., 1974, ApJ, 187, 425
Prugniel P., Simien F., 1997, A&A, 321, 111
Quinlan G. D., Hernquist L., 1997, New Astron., 2, 533
Rajagopal M., Romani R. W., 1995, ApJ, 446, 543
Ravi V., Wyithe J. S. B., Shannon R. M., Hobbs G., 2015, MNRAS, 447, 2772
Reardon D. J. et al., 2016, MNRAS, 455, 1751
Riebe K. et al., 2011, preprint (arXiv:1109.0003)
Roebber E., Holder G., Holz D. E., Warren M., 2016, ApJ, 819, 163
Roedig C., Dotti M., Sesana A., Cuadra J., Colpi M., 2011, MNRAS, 415, 3033
Rosado P. A., Sesana A., Gair J., 2015, MNRAS, 451, 2417
Salucci P., Szuszkiewicz E., Monaco P., Danese L., 1999, MNRAS, 307, 637
Sathyaprakash B. S., Schutz B. F., 2009, Living Rev. Relativ., 12, 2
Savorgnan G. A. D., Graham A. W., Marconi A. r., Sani E., 2016, ApJ, 817, 21
Sazhin M. V., 1978, Sov. Astron., 22, 36
Scannapieco C., White S. D. M., Springel V., Tissera P. B., 2009, MNRAS, 396, 696
Schaye J. et al., 2015, MNRAS, 446, 521
Schmidt M., 1963, Nature, 197, 1040
Schulze A., Wisotzki L., 2010, A&A, 516, A87
Sersic J. L., 1968, Atlas de Galaxias Australes
Sesana A., 2010, ApJ, 719, 851
Sesana A., 2013, MNRAS, 433, L1
Sesana A., Khan F. M., 2015, MNRAS, 454, L66
Sesana A., Haardt F., Madau P., Volonteri M., 2004, ApJ, 611, 623
Sesana A., Haardt F., Madau P., 2006, ApJ, 651, 392
Sesana A., Haardt F., Madau P., 2007, ApJ, 660, 546
Sesana A., Vecchio A., Colacino C. N., 2008, MNRAS, 390, 192
Sesana A., Vecchio A., Volonteri M., 2009, MNRAS, 394, 2255
Sesana A., Barausse E., Dotti M., Rossi E. M., 2014, ApJ, 794, 104
Sesana A., Shankar F., Bernardi M., Sheth R. K., 2016, MNRAS, 463, L6
Shankar F. et al., 2016, MNRAS, 460, 3119
Shankar F. et al., 2019, MNRAS, 485, 1278
Shankar F., Salucci P., Granato G. L., De Zotti G., Danese L., 2004, MNRAS, 354, 1020
Shankar F., Weinberg D. H., Miralda-Escudé J., 2009, ApJ, 690, 20
Shankar F., Weinberg D. H., Miralda-Escudé J., 2013, MNRAS, 428, 421
Shannon R. M. et al., 2015, Science, 349, 1522
Shen X., Hopkins P. F., Faucher-Giguère C.-A., Alexander D. M., Richards G. T., Ross N. P., Hickox R. C., 2020, MNRAS, 495, 3252
Shibuya T., Ouchi M., Harikane Y., 2015, ApJS, 219, 15
Siana B. et al., 2008, ApJ, 675, 49
Siwek M. S., Kelley L. Z., Hernquist L., 2020, MNRAS, 498, 537
Skillman S. W., Warren M. S., Turk M. J., Wechsler R. H., Holz D. E., Sutter P. M., 2014, preprint (arXiv:1407.2600)
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Susobhanan A. et al., 2021, Publ. Astron. Soc. Aust., 38, e017
Taylor J. E., Babul A., 2001, ApJ, 559, 716
Terzić B., Graham A. W., 2005, MNRAS, 362, 197
Tormen G., 1997, MNRAS, 290, 411
Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786, 104
Vasiliev E., Antonini F., Merritt D., 2014, ApJ, 785, 163
Verbiest J. P. W. et al., 2016, MNRAS, 458, 1267
Vestergaard M., Peterson B. M., 2006, ApJ, 641, 689
Vogelsberger M. et al., 2014a, MNRAS, 444, 1518
Vogelsberger M. et al., 2014b, Nature, 509, 177
Volonteri M. et al., 2020, MNRAS, 498, 2219
Volonteri M., Haardt F., Madau P., 2003, ApJ, 582, 559
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Wyithe J. S. B., Loeb A., 2003, ApJ, 590, 691
Yu Q., 2002, MNRAS, 331, 935

This paper has been typeset from a TEX/LATEX file prepared by the author.