

Probing $z \gtrsim$ six massive black holes with gravitational waves

Srija Chakraborty,^{1*} Simona Gallerani,¹ Tommaso Zana^{1D},¹ Alberto Sesana,^{2,3} Milena Valentini,^{4,5,6,7} David Izquierdo-Villalba^{1D},^{2,3} Fabio Di Mascia,¹ Fabio Vito⁸ and Paramita Barai^{1D}⁹

¹*Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa PI, Italy*

²*Dipartimento di Fisica ‘G. Occhialini’, Università degli Studi di Milano-Bicocca, Piazza della Scienza 3, I-20126 Milano, Italy*

³*INFN, Sezione di Milano-Bicocca, Piazza della Scienza 3, 20126 Milano, Italy*

⁴*Universitäts-Sternwarte, Fakultät für Physik, Ludwig-Maximilians Universität München, Scheinerstr. 1, D-81679 München, Germany*

⁵*Excellence Cluster ORIGINS, Boltzmannstr. 2, D-85748 Garching, Germany*

⁶*INAF – Osservatorio Astronomico di Trieste, via Tiepolo 11, I-34131 Trieste, Italy*

⁷*Astronomy Unit, Department of Physics, University of Trieste, via Tiepolo 11, I-34131 Trieste, Italy*

⁸*INAF – Osservatorio di Astrofisica e Scienza dello Spazio Bologna, via Piero Gobetti 93/3, I-40129 Bologna, Italy*

⁹*Centro de Ciências Naturais e Humanas, Universidade Federal do ABC (CCNH-UFABC), Av. dos Estados 5001, Santo André - SP, 09210-580, Brazil*

Accepted 2023 May 13. Received 2023 May 11; in original form 2022 September 26

ABSTRACT

We investigate the coalescence of massive black hole ($M_{\text{BH}} \gtrsim 10^6 M_{\odot}$) binaries (MBHBs) at $6 < z < 10$ by adopting a suite of cosmological hydrodynamical simulations of galaxy formation, zoomed-in on biased ($> 3\sigma$) overdense regions ($M_h \sim 10^{12} M_{\odot}$ dark matter haloes at $z = 6$) of the Universe. We first analyse the impact of different resolutions and AGN feedback prescriptions on the merger rate, assuming instantaneous mergers. Then, we compute the halo bias correction factor due to the overdense simulated region. Our simulations predict merger rates that range between 3 and 15 yr^{-1} at $z \sim 6$, depending on the run considered, and after correcting for a bias factor of $\sim 20\text{--}30$. For our fiducial model, we further consider the effect of delay in the MBHB coalescence due to dynamical friction. We find that 83 per cent of MBHBs will merge within the Hubble time, and 21 per cent within 1 Gyr, namely the age of the Universe at $z > 6$. We finally compute the expected properties of the gravitational wave (GW) signals and find the fraction of LISA detectable events with high signal-to-noise ratio ($\text{SNR} > 5$) to range between 66 per cent and 69 per cent. However, identifying the electro-magnetic counterpart of these events remains challenging due to the poor LISA sky localization that, for the loudest signals ($\mathcal{M}_c \sim 10^6 M_{\odot}$ at $z = 6$), is around 10 deg^2 .

Key words: gravitational waves – galaxies: high-redshift – quasars: supermassive black holes.

1 INTRODUCTION

It is now widely agreed that the centres of nearby galaxies host massive black holes (MBHs, $10^6 \lesssim M_{\text{BH}} \lesssim 10^{10} M_{\odot}$) whose masses correlates with several properties of the host galaxy itself (e.g. Magorrian et al. 1998; Kormendy 2001). Observational pieces of evidence also indicate the existence of bright quasars at $z \gtrsim 6$ (e.g. Fan et al. 2006; Jiang, Jing & Lin 2010) powered by supermassive black holes (SMBHs, $10^8 \lesssim M_{\text{BH}} \lesssim 10^{10} M_{\odot}$) that are accreting close to the Eddington rate. It is still a theoretical challenge to explain how such SMBHs have been assembled within 1 Gyr after the birth of the Universe.

These bright, high- z quasars are expected to be located in massive dark matter (DM) haloes ($M_{\text{halo}} \gtrsim 10^{12} M_{\odot}$). According to the hierarchical structure formation scenario (e.g. White & Rees 1978; Peebles 1980; Blumenthal et al. 1984), such massive DM haloes result from the mergers of smaller haloes that formerly harboured the initial seeds of the SMBHs we observe at $z \sim 6$. Several candidates have been proposed so far as SMBH seeds: (i) *light seeds* ($M_{\text{seed}} \sim 10 - 100 M_{\odot}$), formed as remnants of Pop III stars (Madau,

Pozzetti & Dickinson 1998; Heger et al. 2003; Yoshida, Omukai & Hernquist 2008; Hirano et al. 2015) at $z \sim 20\text{--}30$; (ii) *intermediate seeds* ($M_{\text{seed}} \sim 1000 M_{\odot}$), produced in compact nuclear star clusters as a consequence of runaway stellar mergers at $z \sim 10\text{--}20$ (Davies, Miller & Bellovary 2011; Devecchi et al. 2012; Lupi et al. 2014; Mapelli 2016; Reinoso et al. 2018); and (iii) *heavy seeds* or *direct collapse black holes* (DCBHs, $M_{\text{seed}} \sim 10^4 - 10^6 M_{\odot}$), resulting from the rapid collapse of metal poor gas clouds in $z \gtrsim 10$ atomic cooling haloes (virial temperature $T_{\text{vir}} \geq 10^4 \text{ K}$) where star formation is prevented by intense H_2 photodissociating Lyman Werner (LW) radiation (Haehnelt 1994; Loeb & Rasio 1994; Eisenstein & Loeb 1995; Silk & Rees 1998; Shang, Bryan & Haiman 2010; Johnson et al. 2012; Yue et al. 2013; Ferrara et al. 2014).

Numerous uncertainties remain associated with each of the scenarios discussed above (e.g. Volonteri, Haardt & Madau 2003; Koushiappas, Bullock & Dekel 2004; Begelman, Volonteri & Rees 2006; Lodato & Natarajan 2006; Tanaka & Haiman 2009; Volonteri & Bellovary 2012; Latif et al. 2013; Woods et al. 2019). To understand which of the aforementioned scenarios is the most promising in order to explain the presence of SMBHs at $z \sim 6$, it is necessary to study the still undiscovered population of intermediate-mass black holes (IMBHs, $10^5 \lesssim M_{\text{BH}} \lesssim 10^7 M_{\odot}$) at $6 \lesssim z \lesssim 10$. Various works have investigated the possibility that these IMBHs lie at the centers of

* E-mail: srija.chakraborty@sns.it

dwarf galaxies (Reines, Greene & Geha 2013; Silk 2017; Barai & de Gouveia Dal Pino 2019), and it is difficult to detect them at high- z because of low surface brightness of the galaxies as well as the low intensity of the IMBHs.

The proposed space-borne gravitational wave (GW) observatory Laser Interferometer Space Antennae (LISA; eLISA Consortium et al. 2013), due to be launched in 2034, is potentially an exquisite tool to accomplish this goal. LISA is in fact designed to be sensitive to signals in the frequency range 10^{-4} – 10^{-1} Hz, thus capable to detect GWs from massive (10^4 – $10^8 M_{\odot}$) BH binaries at very-high redshift (even $z \sim 20$ if BHs in this mass range already exist at such early epochs; Haehnelt 1994; Jaffe & Backer 2003; Wyithe & Loeb 2003; Sesana et al. 2004, 2005; Amaro-Seoane et al. 2022). Different scenarios of BH seeds are expected to leave signatures in all those observables that can be probed by GW observations of MBHBs (e.g. merger rates and BH mass distribution; Bhowmick et al. 2022).

Several studies have used semi-analytical models (SAMs) to estimate the MBHB merger rate. For instance, Sesana, Volonteri & Haardt (2007) made predictions for the merger rate of MBHBs when the compact objects start their evolution either as DCBHs or as Pop III star remnants. In this study, the authors considered different stellar feedback regimes to trace the metal enrichment in the gas haloes of MBH formation. Differently, Hartwig, Agarwal & Regan (2018) quantified the rate of *in situ* mergers from binaries of DCBHs. Dayal et al. (2019) investigated the dependence of the merger rate on the seed mass (light versus heavy seeds), the merger prescription (instantaneous versus delayed merger), and cosmic reionization. Barausse et al. (2020) focused on different seeding models, including various astrophysical processes, such as supernovae feedback. Comparably, numerous SAMs, such as Begelman et al. (2006), Arun et al. (2009), Klein et al. (2016), Bonetti et al. (2019), and Valiante et al. (2020) explored different evolutionary channels of MBHB formation (seed mass, accretion efficiency, metal enrichment, and merger time-scales) that could be detected by milli-Hz GW facilities, like LISA.

A different approach instead takes advantage of cosmological hydrodynamical simulations (HDS; e.g. Di Matteo et al. 2012; Vogelsberger et al. 2014; Schaye et al. 2015; Volonteri et al. 2017; Dong-Páez et al. 2023) to study the dynamics of gas and its effect on the evolution of high redshift quasars. Salcido et al. (2016) studied the SMBH mergers occurring in a fully cosmological simulation. In particular, the authors looked for the expected LISA detection rate in the EAGLE run (Schaye et al. 2015), including also a prescription to delay the mergers. More recently, several works made predictions for the LISA detection rate taking advantage of the *Illustris* simulations (Nelson et al. 2015). Particularly, Katz et al. (2019) analysed the effect of different evolutionary models for MBHB and studied the detectability of the merger events (see also Katz & Larson 2019). DeGraf & Sijacki (2019) studied how different seeding models can affect the statistical properties of the MBH population, and the resulting merger rate. Furthermore, DeGraf et al. (2021) explored the relation between SMBH mergers and the morphology of their host galaxies at $z \leq 4$.

In this work, we adopt cosmological zoom-in HDS of galaxy formation, based on the GADGET-3 code, to investigate the coalescence of MBH ($M_{\text{BH}} \gtrsim 10^6 M_{\odot}$) binaries at $6 < z < 10$. We consider a suite of simulations which differ both in terms of resolution and in the stellar/AGN feedback prescriptions implemented and investigate the impact of resolution and feedback on the merger rates of MBHB. For our fiducial model, we further investigate the associated GW properties (chirp mass, merger rate, characteristic strain, signal-to-noise ratio, and angular resolution) of the MBHBs. We also quantify the effect of considering delays in MBHB merging due to dynamical

friction on such GW properties. The paper is organized as follows: in Section 2, we describe the numerical simulations analysed in this work; in Section 3, we study the merger rate resulting from our simulations; and in Section 4, we analyse the effect of adding delays in the MBHB coalescence in post-processing. In Section 5, we analyse the GW properties from the MBHB mergers. Finally, we discuss our results in Section 6 and draw our conclusions in Section 7.

2 SIMULATIONS

In this section, we describe the cosmological HDS adopted in this work. We select simulations from the suites introduced by Valentini, Gallerani & Ferrara (2021) – hereafter V21 – and Barai et al. (2018) – hereafter B18.

The simulations are performed with the TreePM (particle mesh) + SPH (smoothed particles hydrodynamics) code GADGET-3, an evolution of the public GADGET-2 code (Springel 2005), and follow the evolution of a $\sim 10^{12} M_{\odot}$ halo at $z = 6$. In particular, we consider the following runs from V21:

- (i) *AGN_fid*: our fiducial model, featuring thermal AGN feedback;
- (ii) *BHs_noFB*: a control run, analogous to *AGN_fid*, in which BHs and their accretion are included, but AGN feedback is turned off;

and the following runs from B18:

- (i) *AGNcone*: in which the kinetic feedback is distributed in a bi-cone with and half-opening angle of 45° .
- (ii) *AGNsphere*: featuring isotropic, kinetic AGN feedback.

We summarize in the following sections the other main features of the simulations that are relevant for the present study, while we refer to the aforementioned papers for details.

2.1 Valentini et al. (2021) models: *AGN_fid* and *BHs_noFB*

2.1.1 Initial conditions and resolution

The initial conditions are generated with the code MUSIC¹ (Hahn & Abel 2011), assuming a Λ CDM cosmology.² First, a DM-only simulation is run from $z = 100$ to $z = 6$, with DM particles having a mass of $9.4 \times 10^8 M_{\odot}$ in a comoving volume of (148 Mpc).³ Then, a halo as massive as $M_{\text{halo}} = 1.12 \times 10^{12} M_{\odot}$ at $z = 6$ is selected for a zoom-in procedure, to run the full HDS. In the zoom-in region, the highest resolution particles have a mass of $m_{\text{DM}} = 1.55 \times 10^6 M_{\odot}$ and $m_{\text{gas}} = 2.89 \times 10^5 M_{\odot}$. The gravitational softening lengths are³ $\epsilon_{\text{DM}} = 0.72$ ckpc and $\epsilon_{\text{bar}} = 0.41$ ckpc for DM and baryon particles, respectively.

2.1.2 Sub-resolution physics

- (i) Cooling, star formation, and stellar feedback: the multiphase interstellar medium (ISM) is described by means of the MULTI Phase

¹MUSIC – Multiscale Initial Conditions for Cosmological Simulations: <https://bitbucket.org/ohahn/music>.

²We adopt the following parameters by the Planck Collaboration et al. (2016): $\Omega_{\text{M},0} = 0.3089$, $\Omega_{\Lambda,0} = 0.6911$, $\Omega_{\text{B},0} = 0.0486$, and $H_0 = 67.74 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

³We use the following convention when indicating distances: a letter *c* before the corresponding unit refers to *comoving* distances (e.g. ckpc), while the letter *p* refers to *physical* units (e.g. pkpc). When not explicitly stated, we are referring to physical distances.

Particle Integrator (MUPPI) sub-resolution model (Murante et al. 2010, 2015; Valentini et al. 2017, 2019). It features metal lines cooling, an H₂-based star formation, thermal and kinetic stellar feedback, the presence of an UV background, and the Tornatore et al. (2007) model for chemical evolution.

(ii) Black holes seeding and merging: BHs are treated as collisionless sink particles. Seeds of mass $M_{\text{BH, seed}} = 1.48 \times 10^5 M_{\odot}$ are implanted in DM haloes with mass exceeding $M_{\text{DM, seed}} = 1.48 \times 10^9 M_{\odot}$. This seeding prescription is meant to mimic in a simplistic way the DCBH scenario described in the section. Two BHs are allowed to merge when their relative distance becomes smaller than twice the BH gravitational softening length, and their relative velocity is lower than the sound speed of the local ISM. The final BH is set on the position of the most massive BH, which underwent the merger. BH repositioning (or *pinning*) is implemented, in order to prevent BHs from wandering from the centre of the halo in which they reside: at each time-step BHs are shifted towards the position of minimum gravitational potential within their softening length (as also done in e.g. Booth & Schaye 2009; Schaye et al. 2015; Weinberger et al. 2017; Pillepich et al. 2018).

(iii) Gas accretion on BHs: besides BH–BH mergers, black holes are also allowed to grow via gas accretion, as described by the classical Bondi–Hoyle–Lyttleton (BHL) model (Hoyle & Lyttleton 1939; Bondi & Hoyle 1944; Bondi 1952; Edgar 2004):

$$\dot{M}_{\text{Bondi}} = \frac{4\pi G^2 M_{\text{BH}}^2 \rho}{(c_s^2 + v^2)^{3/2}}, \quad (1)$$

where G is the gravitational constant, M_{BH} is the BH mass, ρ is the gas density, c_s is the sound speed, and v is the velocity of the BH relative to the gas. These quantities are evaluated by averaging over the SPH gas particles within the BH smoothing length, with kernel-weighted contributions. Equation (1) is used to estimate the contribution to the accretion rate from the cold and hot phase of the ISM, separately (Steinborn et al. 2015; Valentini et al. 2020). Accretion from the cold gas is reduced by taking into account its angular momentum (see Valentini et al. 2020, for details). The BH accretion rate is capped to the Eddington accretion rate.

(iv) Quasar feedback: a fraction of the accreted rest-mass energy is radiated away with a radiative efficiency ϵ_r , thereby providing a bolometric luminosity for an accreting BH equals to:

$$L_{\text{bol}} = \epsilon_r \dot{M}_{\text{BH}} c^2, \quad (2)$$

where c is the speed of light and $\epsilon_r = 0.03$ (Sądowski & Gaspari 2017). Then, a fraction $\epsilon_f = 10^{-4}$ (V21) of the radiated luminosity L_{bol} is coupled thermally and isotropically to the gas surrounding the BH. The AGN feedback energy is distributed to the hot and cold phases of the multiphase gas particles within the BH smoothing volume (Valentini et al. 2020).

2.2 Barai et al. (2018) simulations: *AGNcone* and *AGNsphere*

2.2.1 Initial conditions and resolution

Initial conditions are generated as in V21 with the code MUSIC and adopting the same cosmology. The parent, DM-only simulation follows a comoving volume of $(500 \text{ Mpc})^3$, with DM particles having $m_{\text{DM}} = 2 \times 10^{10} M_{\odot}$. The zoom-in run focuses on a DM halo as massive as $M_{\text{halo}} = 4.4 \times 10^{12} M_{\odot}$; the highest resolution particles have $m_{\text{DM}} = 7.54 \times 10^6 M_{\odot}$ and $m_{\text{gas}} = 1.41 \times 10^6 M_{\odot}$, with a gravitational softening length $\epsilon_{\text{bar}} = \epsilon_{\text{DM}} = 1.48 \text{ ckpc}$.

2.2.2 Sub-resolution physics

(i) Cooling, star formation, and stellar feedback: radiative heating and cooling is accounted for by employing the CLOUDY cooling tables computed by Wiersma, Schaye & Smith (2009). Star formation is implemented following the ISM multiphase model by Springel & Hernquist (2003), in which an hot and a cold phase co-exist in pressure equilibrium, and assuming a density threshold for the star formation of $n_{\text{SF}} = 0.13 \text{ cm}^{-3}$. A Chabrier (2003) initial mass function (IMF) in the mass range $(0.1–100) M_{\odot}$ is adopted. Stellar evolution and chemical enrichment are computed following Tornatore et al. (2007).

(ii) Black hole seeding and merging: As in B18, the theoretical mass of seed BHs is $10^5 M_{\odot}$. However, their dynamical mass is much smaller in the V21 simulations ($\sim 10^5 m_{\odot}$ in V21 versus $\sim 10^7 M_{\odot}$ in B18). The prescription is in fact more refined in V21, where seeded BHs are linked to stellar particles instead of DM particles. We further discuss this point in Section 3.1.

(iii) Gas accretion on BHs: as in V21, the BHL model is adopted. However, the lower resolution of the B18 simulations does not allow to properly describe the accretion process: thus equation (1) is multiplied by a numerical boost factor $\alpha = 100$ (Springel, Di Matteo & Hernquist 2005; Sijacki, Springel & Haehnelt 2009; Vogelsberger et al. 2014). No angular momentum effects are included in the B18 formalism, and no distinction between the hot and cold gas phases is considered.

(iv) Quasar feedback: black holes are assumed to radiate energy away with an efficiency of $\epsilon_r = 0.1$, and a fraction $\epsilon_f = 0.05$ of this energy is coupled to the surrounding gas via kinetic feedback as an energy-driven wind (see Barai et al. (2018) for details). The geometry of the feedback is bi-conical (i.e. energy is injected on to a bi-cone with a half-opening angle of 45°) in *AGNcone* and spherical (i.e. energy distributed isotropically) in *AGNsphere*. BHs grow ~ 10 times more massive at $z = 6$ in the *AGNcone* case than in the *AGNsphere* run as shown in top panels of fig. 2 in B18. This is because in the *AGNcone* run more gas can inflow along the perpendicular direction to the bi-cone, and accrete on to the black hole.

A summary of the two different simulation models are provided in Table 1 for a comprehensive view.

3 MERGER RATE

3.1 Merger rate from overdense regions

For the models summarized in Table 1, we compute the redshift evolution of the MBH merger rate (per unit redshift, per unit time) for different chirp mass ranges, where the chirp mass is given by (e.g. Cutler & Flanagan 1994; Blanchet et al. 1995):

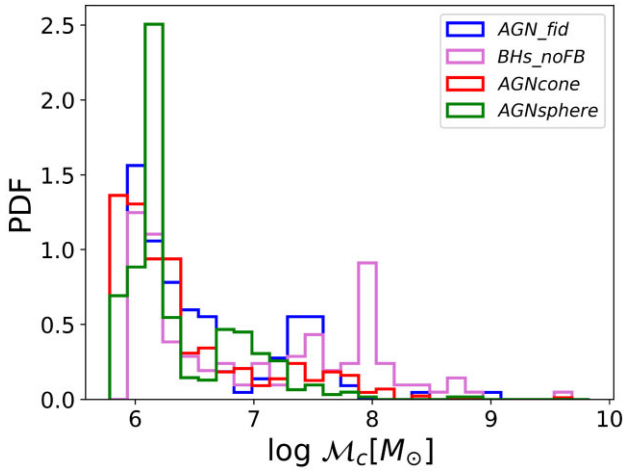
$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}, \quad (3)$$

and m_1 and m_2 are the masses of the merging black holes.⁴ We show in Fig. 1 the normalized probability distribution function (PDF) of the chirp masses resulting from different simulations. Although no evident differences among different models can be seen from this plot, we note that the B18 simulations predict a larger number of small chirp masses ($\mathcal{M}_c < 10^6 M_{\odot}$) with respect to V21, and

⁴In these calculations, we do not include those MBHBs that cannot be associated with any galaxy in the simulations. We discuss these spurious events in Section 4.1.

Table 1. Summary of the main features of the suites of the cosmological HDS adopted in this work (Valentini et al. 2021; Barai et al. 2018).

	<i>AGN_fid</i>	<i>BHs_noFB</i>	<i>AGNcone</i>	<i>AGNsphere</i>
$z = 6$	Mass resolution [M_\odot]: $m_{\text{DM}} = 1.5 \times 10^6$ $m_{\text{gas}} = 2.9 \times 10^5$ Gas particle smoothing length[pc] = 59 Size of the zoomed region = 5.25cMpc DM halo host: $M_{\text{halo}} = 1.1 \times 10^{12} M_\odot$ $\epsilon_r = 0.03$ V21		Mass resolution [M_\odot]: $m_{\text{DM}} = 7.54 \times 10^6$ $m_{\text{gas}} = 1.41 \times 10^6$ Gas particle smoothing length[pc] = 211 Size of the zoomed region = 5.21cMpc DM halo host: $M_{\text{halo}} = 4.4 \times 10^{12} M_\odot$ $\epsilon_r = 0.1$ B18	
M_\star [M_\odot]	4×10^{10}	3.5×10^{10}	7×10^{10}	6×10^{10}
SFR [$M_\odot \text{yr}^{-1}$]	200	190	200	300
M_{BH} [M_\odot]	10^9	5×10^{11}	2×10^9	5×10^8
BHAR [$M_\odot \text{yr}^{-1}$]	35	3×10^4	89	3
Feedback	Stellar, AGN (thermal)	Stellar	Stellar, AGN (kinetic, bi-conical geometry)	Stellar, AGN (kinetic, spherical geometry)


Figure 1. PDF of chirp mass as resulting from the *AGN_fid* (blue), *BHs_noFB* (magenta), *AGNcone* (red), and *AGNsphere* (green).

only the *BHs_noFB* and *AGNcone* simulations predict MBHBs with $M_c > 10^9 M_\odot$.

We then calculate the merger rate from the number density of mergers, dN/dz per comoving volume, dV as (e.g. Haehnelt 1994; Ciardi & Loeb 2000):

$$\frac{dN}{dz dt_{\text{obs}}} = \frac{4\pi c d_L^2}{(1+z)^2} \frac{dN}{dz dV}, \quad (4)$$

where d_L is the luminosity distance of the event. The results are shown in Fig. 2.

The number of MBH mergers predicted by our zoom-in cosmological simulations increases with decreasing redshift, as a consequence of the hierarchical structure formation process. Fig. 2 shows that the two sets of simulations have different merging histories that depend both on the numerical resolution adopted and on the feedback implemented.

In particular, Fig. 2 shows that the number of mergers predicted by the V21 simulations (~ 100 at $z = 6.5$ for both *AGN_fid* and *BHs_noFB*) is smaller by a factor of at least 2 than the one by the B18 simulations (> 200 for both *AGNcone* and *AGNsphere* case at $z \sim 6.5$). This major difference is likely linked to the better numerical resolution of the V21 simulations with respect to the B18 ones (see Table 1). The condition for seeding ($M_h > 10^9 M_\odot$) is more easily

satisfied in B18 with respect to V21, because of the larger mass of its resolution elements. This thus implies a larger number of seeds and consequently a larger number of mergers. Even more importantly, the excess of mergers in B18 is driven by the high dynamical mass, which enters in the BH–BH merger algorithm within the code.

For a fixed numerical resolution, it is possible to study the dependence of the number of mergers on the feedback implemented. In the V21 simulations, we find that for low chirp masses ($< 10^7 M_\odot$) the merger rate predicted by the *AGN_fid* model is higher than the *BHs_noFB* case. We further note that for $M_c > 10^7 M_\odot$ this trend is reversed, the merger rate in the *BHs_noFB* run is higher than in the *AGN_fid* case. The lack of AGN feedback allows a more efficient gas accretion, which makes the black holes to grow more massive and numerous in the *BHs_noFB* case in this particular chirp mass range.

Furthermore, we study the trend of the local sound speed in the ISM. In the left-hand panels of Fig. 3 we show the results for the V21 simulation runs. We note that the sound speed is larger in the *BHs_noFB* run because of the following. In the *BHs_noFB* run, the accretion rate is higher than in *AGN_fid* (see Table 1); furthermore, on scales over which c_s is computed (i.e. the smoothing length of the black holes), the heating due to accretion (gravitational compression) dominates the heating due to feedback. Thus, the gas temperature (and consequently c_s) in the *BHs_noFB* run is higher than *AGN_fid*. We, thus conclude that for $M_c < 10^7 M_\odot$, the merger rate is driven mainly by the gravity, dynamics, and substructure mergers.

For the B18 runs, we observe from Fig. 2 that in *AGNcone* the number of mergers is larger by a factor of ~ 2 with respect to the *AGNsphere* case. To investigate this point, in the right-hand panels of Fig. 3 we show the PDF of the relative sound speed of the merging BHs resulting from these simulations. This figure shows that in the *AGNcone* case, the PDF is shifted towards larger values. As discussed in Section 2, two BHs are allowed to merge when their relative velocity is lower than the sound speed of the local ISM. The higher is the sound speed the larger is the probability for two BHs to merge. The sound speed is larger in the *AGNcone* run because in this case the accretion is by far higher than in *AGNsphere* (see Table 1), resulting in a higher number of mergers.

We further note that Zana et al. (2022) already found that different feedback prescriptions result into different merger rates: in *AGNcone*, galaxies merge faster and more easily than in *AGNsphere*, possibly because of the stronger feedback due to the larger black hole accretion rate (see Table 1). This determines a more diffuse gas and stellar component around the host galaxies, which can boost the effect of

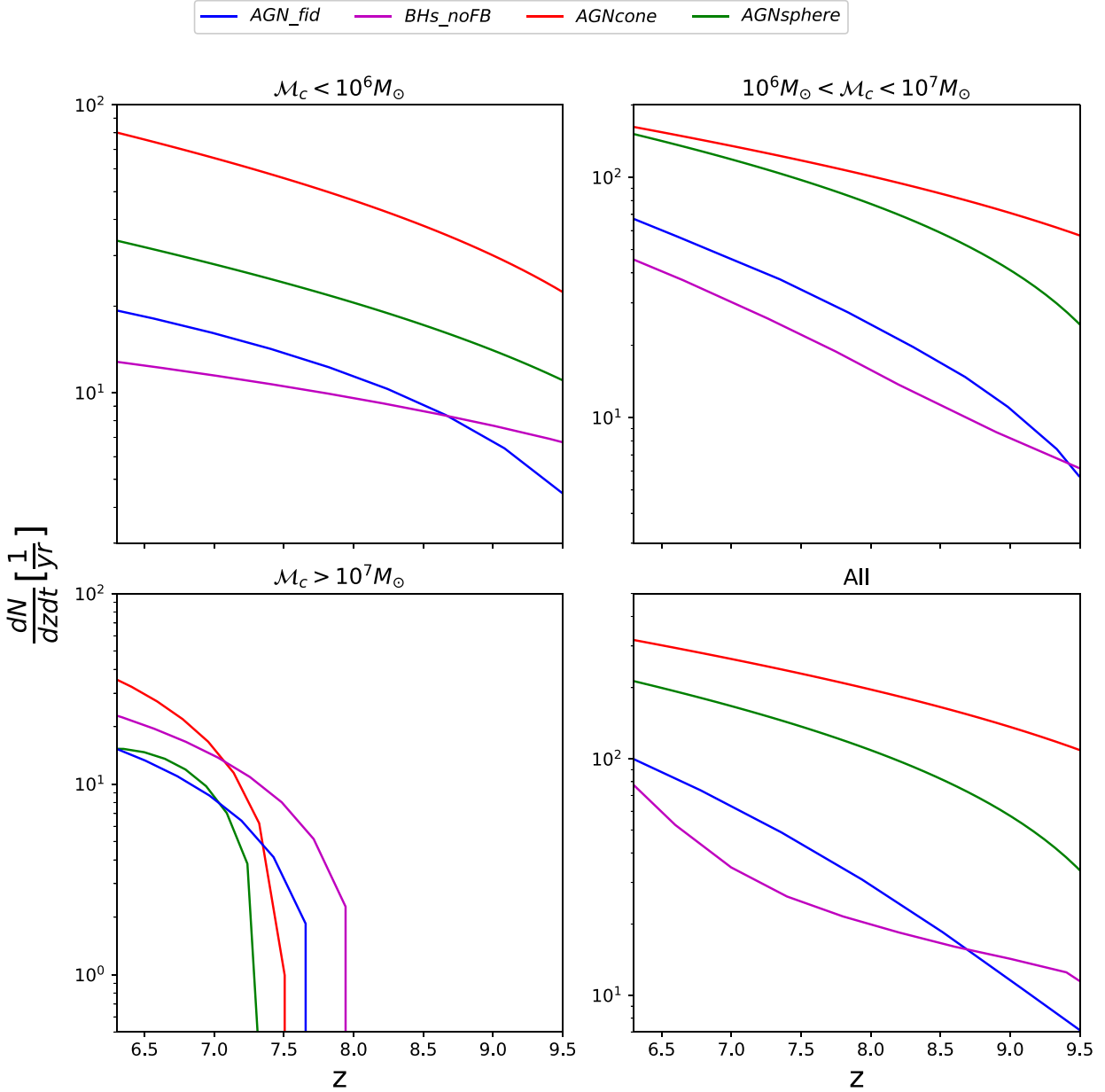


Figure 2. Number of mergers per unit redshift per unit time for *AGN_fid* (blue), *BHs_noFB* (purple), *AGNcone* feedback (red), and *AGNsphere* feedback (green). The upper left-hand (right) panel shows the merger rates for MBHB systems with chirp mass $< 10^6 M_\odot$ ($10^6 < \mathcal{M}_c < 10^7 M_\odot$); the lower left-hand panel shows the merger rates for chirp mass $> 10^7 M_\odot$, while the lower right-hand panel refers to the cumulative merger rates for all chirp mass ranges.

dynamical friction when two galaxies approach, thus lowering the dynamical time-scale for their merging to occur.

We finally note that different resolutions and feedback prescriptions also affect the epoch at which the furthest merger event is occurring. For example, for $\mathcal{M}_c > 10^7 M_\odot$, the furthest GW signal occurs in the redshift interval $7.7 < z < 8$ and $7.3 < z < 7.5$ in the case of V21 and B18 simulations, respectively.

3.2 Bias in zoom-in simulations

Our results are based on small box simulations, zoomed-in on a massive, biased ($> 3\sigma$; e.g. Barkana & Loeb 2001) DM halo. As a result, the number density of MBHB mergers within the simulated box overestimates the value in an average region of the Universe. In

what follows, we estimate the bias of our predictions considering our fiducial run (*AGN_fid*). Following the halo mass history suggested by Correa et al. (2015a), we first calculate the accreted mass ($M_{h,0}$) at $z = 0$ of a $10^{12} M_\odot$ halo at $z = 6$:

$$M_h(z) = M_{h,0} (1+z)^\alpha e^{\beta z}, \quad (5)$$

where $\alpha = 0.24$ and $\beta = -0.75$, respectively (Correa et al. 2015b). We find $M_{h,0} = 10^{13.75} M_\odot$.

In practice, when computing the merger rate with equation (4), we are considering a Universe where all DM haloes have $M_h = 10^{13.75} M_\odot$ and comoving volume $(5.25 \text{ cMpc})^3$. The result of this procedure will be clearly biased compared to a proper calculation in which DM haloes span a wider mass range ($M_h = 10^{10} \sim 10^{16} M_\odot$) and have different abundances. This bias cannot be directly computed

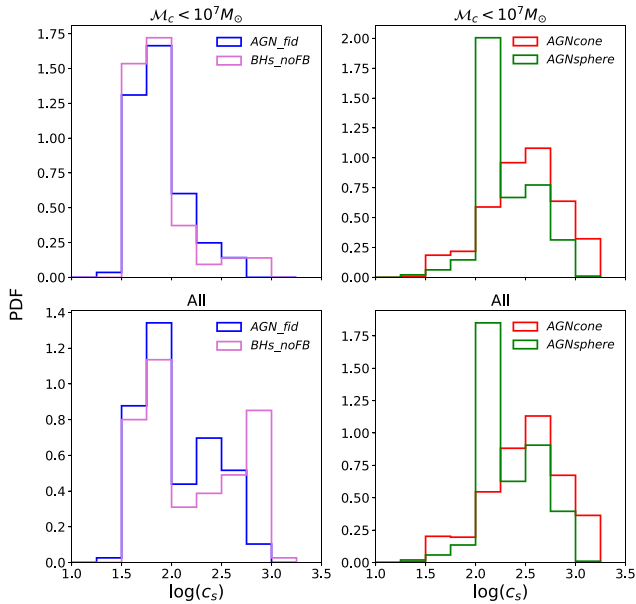


Figure 3. PDF of relative sound speed of merging black holes for *AGN_fid* feedback and *BHS_noFB* from V21 (left-hand panel) and *AGNcone* and *AGNsphere* feedback from B18 (right-hand panel). The top panels show the distribution of the sound speeds of MBHBs with $\mathcal{M}_c < 10^7 M_\odot$ while the bottom panels show the same for all MBHBs.

from our simulations, but we need to rely on a SAM. In SAMs, in fact, a wide range of haloes are simulated and the merger rate is computed from their collective output, by weighting each halo mass according to the Press and Schechter halo mass function (Press & Schechter 1974). This latter step is simply obtained by dividing the merger rate by the effective comoving volume occupied by that halo. It is therefore also possible to use the SAM output to create biased universes, simply by taking haloes of a desired mass and weighting them with a desired effective comoving volume. By comparing the rates obtained from this biased universe to the total one, we can infer the bias. This is the procedure we follow here.

We then use the SAM by Barausse (2012; see also Klein et al. 2016), and we consider the output of all the trees, weighted on the Press and Schechter halo mass function (Press & Schechter 1974). For these calculations, we consider the ‘Q3nod’ run, since it is based on a model that more closely resembles our prescriptions (heavy seeds and no time delays). Furthermore, for a fair comparison with our results, we only consider $z > 6$ merger events occurring in DM haloes $> 10^9 M_\odot$ (the threshold mass used in our simulations to seed MBHBs) and involving binaries with both BH masses $> 10^5 M_\odot$ (the mass of our seeds). The merger rate obtained in this case is $\sim 3 \text{ yr}^{-1}$.

We next consider two specific merger trees, whose haloes at $z = 0$ are the closest to the $M_{h,0}$ value computed above, namely $M_{h1,0} = 10^{13.7} M_\odot$ and $M_{h2,0} = 10^{13.8} M_\odot$. We weight the merger rates in these two merger trees with the inverse of the comoving volume of our refined simulation, namely $(5.25 \text{ cMpc})^3$. The average of the two merger rates obtained in this way can be used as a proxy for the merger rate in a Universe made only of haloes of $M_{h,0} = 10^{13.75} M_\odot$ at $z = 0$. The merger rate obtained in this case is $\sim 60 \text{ yr}^{-1}$.

We, thus estimate that the merger rates computed from our zoom-in simulations is biased by a factor ~ 20 . We repeat the above calculations for the model ‘popIII’ (which assumes light seeds and delays between MBH and galaxy mergers) and the model ‘Q3-d’

Table 2. Comparison with contemporary literature for merger rates at $z = 6$. The results of our work are reported after being corrected for the halo bias computed in Section 3.2.

References	Model	$\frac{dN}{dz dt} [\text{yr}^{-1}]$
Sesana et al. (2007)	SAM	1.5–25
Klein et al. (2016)	SAM	10
Hartwig et al. (2018)	SAM	5
Dayal et al. (2019)	SAM	< 10
Katz et al. (2019)	HDS	0.01
This work (bias corrected)	HDS	3–15

(heavy MBH seeds and delays). In both cases the bias does not change significantly, being ~ 20 – 30 .

To summarize, our zoom-in simulations predicts a total number of merger events per year, at $z \sim 6$, that varies between 80 and 300, depending on the resolution and the star formation/AGN feedback prescriptions adopted. By accounting for the halo bias, the number of merger events lower to ~ 3 – 15 yr^{-1} .

3.3 Comparison with contemporary works

In this section, we compare our results at $z = 6$ with contemporary works that make similar calculations, both using SAMs and HDS. For this comparison, we only consider those models that do not include any time delay in MBHB coalescence, which is consistent with our work.⁵ We summarize results from different models in Table 2.

We first compare our results with the predictions by Sesana et al. (2007) based on the models by Begelman et al. (2006). In these works, DCBH formation is efficient when the haloes overcome a given threshold of virial temperature ($T_{\text{vir}} \geq 10^4 \text{ K}$): the ‘high-feedback’ (BVRhf) and ‘low-feedback’ (BVRlf) models differ for the efficiency in the distribution of metals produced during the star formation process. In the BVRhf (BVRlf) model, the merger rate is 1.5 (25) yr^{-1} . This difference is due to the following: in BVRhf, as a consequence of the high stellar feedback efficiency that ensures a swift metal enrichment, the DCBH formation stops as early as $z \sim 18$; in the BVRlf, the DCBH formation only stops at $z \sim 15$, since the low stellar feedback efficiency allows haloes to remain pristine longer.

Klein et al. (2016) also investigated the effect of different seeding models and time delays between galactic and MBHB mergers on the merger rate. For this, they adopted the SAM by Barausse (2012) and varied the seeding mass (light seed versus heavy seed), and hence their halo occupation fraction. They also consider delays in MBHB caused by MBH environment as well as by triple interactions. Considering the model with heavy seeds and instantaneous mergers (model ‘Q3-nod’) at redshift 6, they predict a merger rate of 10 per year.

We further consider the results by Hartwig et al. (2018) for instantaneously merging binaries with $10^4 < M_{\text{seed}} < 10^6 M_\odot$, as derived by assuming a critical LW flux⁶ of $J_c = 30 J_{21}$. In the Hartwig et al. (2018) calculations, the merger rate the DCBH formation rate peaks at approximately $z \sim 7$ after which the DCBH formation stops, as a result of the metal enrichment and cosmic reionization processes. The resulting merger rate at $z = 6$ is 10 yr^{-1} . For what concerns Dayal et al. (2019), they use the SAM of galaxy formation *Delphi* to track the effect of different BH parameters on the BH merger rates. The

⁵We further discuss this point in Section 4.

⁶ J_{21} is the LW flux in units of $10^{-21} \text{ erg cm}^{-2} \text{ s}^{-1} \text{ Hz}^{-1} \text{ sr}^{-1}$.

merger rate at $z = 6$ is found to be < 10 per year. We only consider the case of instantaneous mergers which does not assume any delays between galactic and MBH mergers, referred to as *ins1* model in the Dayal et al. (2019). We also note that they only report the intrinsic merger rate for all BH mergers (stellar BBH mergers, ‘mixed’ merger with stellar seed and DCBH as well as DCBH–DCBH mergers). For this reason,⁷ we consider this estimate as an upper limit to the merger rate when only DCBH are seeded.

For what concerns HDS, Katz et al. (2019) used *Illustris* to study different populations of MBHs. They predict the effect on LISA detection rate for different MBH evolutionary scenario such as the effect of delay on the BH particle mergers in the simulations combined with different BH masses. At $z \sim 6$, they predict the intrinsic merger rate ~ 0.01 per year for heavy seeds and instantaneous mergers (model *ND*).

To summarize, our results are consistent with previous predictions from semi-analytical works (after being corrected for the halo-bias), while they are above the predictions by Katz et al. (2019). This inconsistency can be ascribed to the different seeding mechanisms adopted in their work $M_{\text{seed}} = 1.42 \times 10^5 M_{\odot}$ BHs seeded in $M_{\text{DM}} = 7.1 \times 10^{10} M_{\odot}$ DM haloes): since the DM haloes in which BHs are seeded are more massive (i.e. less numerous) than ours, we expect fewer BHs to be seeded in the simulations, resulting into a lower merger rate.

Furthermore, in Appendix A, we compare the gravitational wave background (GWB) resulting from our simulations with current NanoGrav, EPTA, and PPTA observations. We notice that this is only a sanity check to make sure that our predictions are not overshooting current observational constraints. However, it does not represent a genuine comparison since our predictions do not include any contribution from sources at $z < 6$, which are instead expected to dominate the background (e.g. Izquierdo-Villalba et al. 2021).

4 DELAYS IN MBHB MERGERS

In our simulations, given the limited spatial and temporal resolution, we are not able to properly follow the dynamics of MBHBs up to coalescence. This explains the simplified prescription typically adopted in zoom-in cosmological simulations for BH merging described in Section 2. However, the actual time-scale over which MBHBs merge depends on several factors, e.g. the mass ratio of the MBHBs, their initial separation, and the physical properties of the galaxy hosting the MBHB. In what follows, for our fiducial model, we first describe how we associate a MBHB to its host galaxy, and then we correct in post-processing the coalescing time of MBHB mergers including a time delay due to dynamical friction from the surrounding stars.⁸ This allows us to get a first order estimate of how our merger rates change due to these effects. The final results may anyway vary to some degree from what we report below, if the dynamical friction were actually implemented in the code, instead of applying its effect in post-processing.

4.1 Galaxy-MBHB association

In this subsection, we describe the method we adopt to associate a MBHB to its host galaxy. Galaxy identification follows a similar

⁷They also show that DCBH mergers, noted as ‘type 3’ mergers, are the rarest in their BH population.

⁸In Appendix B, we describe the stellar hardening physical process that could further delay the coalescence of MBHBs. We find that the resolution of our simulations prevents us to make realistic predictions about this effect.

approach to what has been done in Zana et al. (2022). We identify⁹ DM haloes through the AMIGA halo finder code (Knollmann & Knebe 2009). The merger tree for each halo at $z \simeq 6$ is built by tracing back in time the constituent DM particles: their ID is matched in the progenitor structures in the previous snapshots. Baryon particles are assigned to their related galaxy when: (i) they are located within βr_{vir} of a given halo, where r_{vir} is the virial radius of the halo and $\beta = 0.3$; and (ii) their velocity is lower than the escape velocity, as evaluated through an analytical integration of the Navarro–Frank–White profile (Navarro, Frenk & White 1996) to speed up calculations. We restrict our analysis only to those galaxies with $M_{\text{vir}} > 10^9 M_{\odot}$ and $M_{*} > 10^7 M_{\odot}$.

We associate a host galaxy to each merger event, by using the following procedure: (i) we first assign to each merger event that galaxy for which its centre of mass is the closest to the position of the primary BH (BH_p); (ii) we consider only those mergers, which are within βr_{vir} of any galaxy. When we associate the galactic properties (derived from a snapshot) to a merger event, we consider the closest¹⁰ snapshot in time to the event itself (see Section 6 for further discussion).

In some cases, our algorithm fails to associate a host galaxy to a merger event. This may occur because during the time passing between the redshift of the event and the closest snapshot, the BH_p may have moved out of the r_{vir} of the host galaxy. Another possibility is related to the fact that during the simulation a MBH may be spuriously seeded into a transient matter overdensity (incorrectly identified as a galaxy by the on-the-fly halo finder); such a MBH would then rapidly merge with the MBH of the closest halo (due to the repositioning algorithm) in less than 1 time-step (e.g. Blecha et al. 2015; Kelley, Blecha & Hernquist 2016; Katz et al. 2019).

We find, in our fiducial model, over the 145 total events, 72 per cent occur within βr_{vir} , 22 per cent outside βr_{vir} but inside r_{vir} , and 6 per cent outside the r_{vir} . We remove these spurious events from our calculations.¹¹

4.2 Time delay due to dynamical friction

The interaction of the MBHBs with stars in their surroundings results in the MBHBs to lose energy, to slow down, and to spiral inwards gradually. (Chandrasekhar 1943; Ostriker 1999). This process effectively increases the time-scale of the MBHB merger with respect to the adopted simulations, potentially delaying it by millions or even billions of years. The amount of dynamical friction experienced by MBHBs depends on the density and distribution of the surrounding stars, as well as the mass and velocity of the binary. In general, the effect of dynamical friction is strongest in regions of high density, such as the centers of galaxies, where the density of DM and stars is the highest. We make a simple calculation following the prescription of Krolik et al. (2019), but see also Volonteri et al. (2020) for more details.

⁹To define a halo we require a minimum of 20 bound particles.

¹⁰We have also considered the case in which galactic properties are linearly interpolated from the two snapshots immediately before and after the merger event, finding no appreciable differences in the main results of our work.

¹¹We also calculate the spurious events in *AGNcone* for comparison of different resolutions and we find that the fraction of spurious events increase with the decrease of resolution of the numerical simulation. Over the 1812 total events in *AGNcone*, only 10 per cent occur within βr_{vir} , 20 per cent outside βr_{vir} but inside r_{vir} , and 70 per cent outside the r_{vir} . This further strengthens our selection of *AGNfid* as our fiducial model.

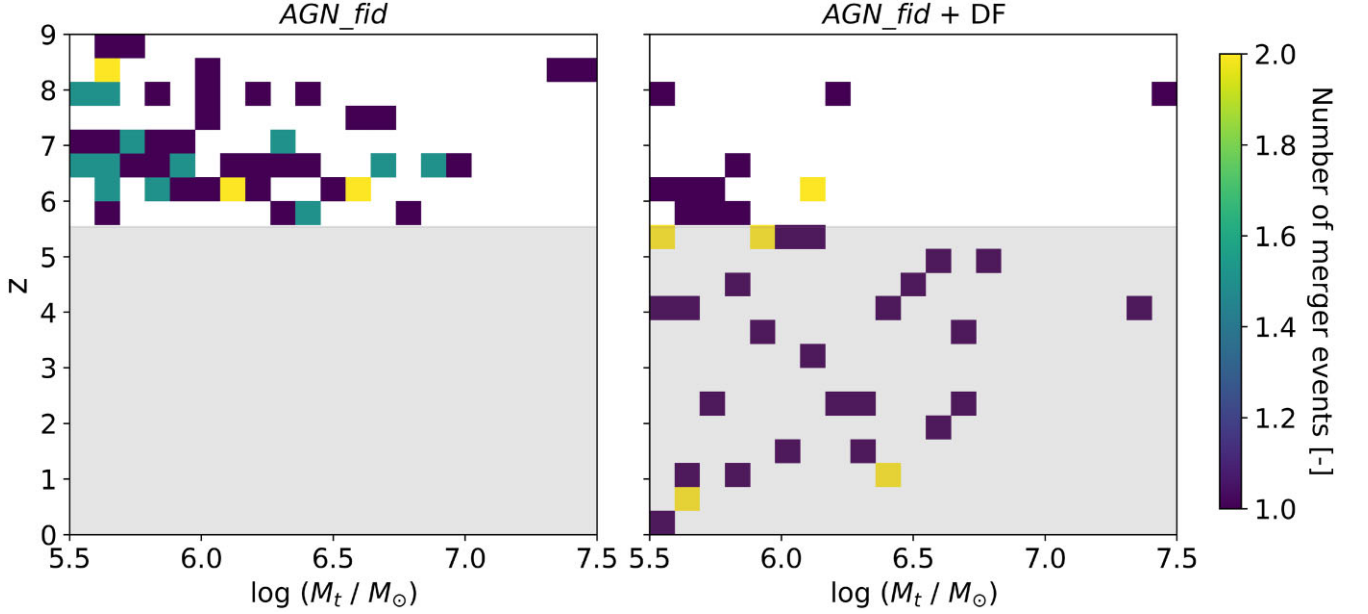


Figure 4. Number of merger events as a function of the total mass of the system M_t and the redshift z of the binary, as resulting from our fiducial run. The left-hand panel shows the merger distribution resulting from the simulations while the right-hand panel shows the same distribution after adding the delay time due to dynamical friction. The white region in each plot shows the redshift range accessible to the simulations ($z > 6$) and the grey region denotes the redshift range inaccessible to our simulations.

The frictional time-scale for a massive object in an isothermal sphere can be written as (Binney & Tremaine 2008):

$$t_{\text{df}} = 0.67 \text{ Gyr} \left(\frac{a}{4 \text{ kpc}} \right)^2 \left(\frac{\sigma}{100 \text{ km s}^{-1}} \right) \left(\frac{M_{\text{BH}_s}}{10^8 M_\odot} \right)^{-1} \frac{1}{\Lambda}, \quad (6)$$

where a is the distance of the MBH from the Galaxy Centre¹², σ is the central stellar velocity dispersion:

$$\sigma = (0.25 G M_* / R_{\text{eff}})^{1/2}, \quad (7)$$

M_* is the total stellar mass of the galaxy hosting the MBHs, computed as described in Section 4.1,

$$\Lambda = \ln(1 + M_* / M_{\text{BH}_s}), \quad (8)$$

$R_{\text{eff}} = 0.1 r_{\text{vir}}$, and M_{BH_s} denotes the mass of the secondary (less massive) MBH.

The total time taken by the MBHBs to merge including the dynamical friction correction is then given by

$$t_{\text{tot,df}} = t_{\text{in}} + t_{\text{df}}, \quad (9)$$

where t_{in} is the time at which the merger occurs in the simulation.

The results of our calculations are reported in Fig. 4, where we show the number of merger events (for which we can associate a host galaxy) across different mass and redshift ranges without including any post-processing delay (left-hand panel), and including dynamical friction (right-hand panel). The grey region in each panel denotes the redshift range outside the reach of our simulations ($z < 6$). Fig. 5 helps to better visualize the difference in the merger rate predictions if we assume instantaneous merger (blue line), or we include delay due to dynamical friction (yellow line) in post-processing. By adding the delay due to dynamical friction, we find that 17 per cent of the MBHBs of our fiducial calculations are not

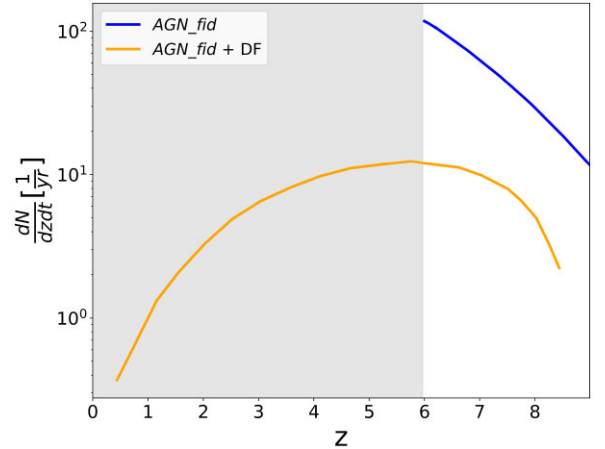


Figure 5. Merger rate per unit redshift z of MBHBs, considering delay times in post-processing. In blue, mergers from *AGN_fid* are shown and the same events, which can be associated with a host galaxy are shown in orange after adding the delay time due to dynamical friction. The grey shaded region depicts the redshift range, which is inaccessible in our simulations.

merging within the Hubble time, 21 per cent of the MBHBs merge at $z > 6$, and the rest will be delayed to a redshift range $z < 6$.

5 GWS FROM HIGH REDSHIFT MBHBS

In this section, we estimate the detectability of the merger events predicted by the V21 *AGN_fid* simulations,¹³ by computing the signal-to-noise ratio (SNR) and the angular resolution Ω of their GW signals. We also show how these properties vary if time delay

¹²We calculate a at the snapshot closest in time to the numerical merger.

¹³In Appendix C, we compare the GW properties reported in this section for the *AGN_fid* with the other simulation runs presented in Section 2.

due to dynamical friction is considered. We assume that a GW signal is detectable if $\text{SNR} > 5$. Hereafter, for the sake of brevity, we refer to these ‘LISA detectable events’ as LDEs. Furthermore, we also depict the fraction of LDEs in the ‘mass ratio’–‘total mass’ plane with and without considering delay effects. Finally, we discuss the different scales of interest in time, frequency and spatial ranges of the LDEs.

5.1 SNR ratio and angular resolution

The SNR accumulated over the observational time τ is computed following the Flanagan & Hughes (1998) formalism:

$$\left(\frac{S}{N}\right)_{\Delta f}^2 = \int_f^{f+\Delta f} d \ln f' \left[\frac{h_c(f')}{h_{\text{rms}}(f')} \right]^2, \quad (10)$$

where f_r is the GW rest-frame frequency, $f = f_r/(1+z)$ is the observed frequency, Δf is the frequency shift in the duration of τ , h_c is the *characteristic strain*, and h_{rms} is the effective¹⁴ rms noise of the instrument.

We start defining the strain amplitude (sky and polarization averaged) of GWs emitted by two black holes of chirp mass \mathcal{M}_c that are merging at redshift z , following Hawking & Israel (1989):

$$h = \frac{8\pi^{2/3}}{10^{1/2}} \frac{G^{5/3} \mathcal{M}_c^{5/3}}{c^4 r(z)} f_r^{2/3}, \quad (11)$$

where $r(z)$ is the luminosity distance of the merging events. MBHBs spend a mass-dependent amount of time in each frequency band, as shown in Fig. 6.

It is, thus common to compute the characteristic strain amplitude that also depends on the number of cycles spent in the LISA bandwidth by the binaries:

$$h_c = h\sqrt{n} \simeq \frac{1}{3^{1/2}\pi^{2/3}} \frac{G^{5/6} \mathcal{M}_c^{5/6}}{c^{3/2} d_L} f_r^{-1/6}, \quad (12)$$

where n is the number of cycles spent in a frequency interval Δf :

$$n \simeq f_r^2 / \dot{f}_r = \frac{5}{96\pi^{8/3}} \frac{c^5}{G^{5/3} \mathcal{M}_c^{5/3}} f_r^{-5/3}, \quad (13)$$

and the rest-frame frequency shift rate is expressed as

$$\dot{f}_r = \frac{df_r}{dt_r} = \frac{96\pi^{8/3} G^{5/3}}{5c^5} \mathcal{M}_c^{5/3} f_r^{11/3}, \quad (14)$$

assuming that the backreaction from GW emission dominates the orbital decay of a binary.

The LISA rms noise h_{rms} is instead given by

$$h_{\text{rms}}(f) \equiv \sqrt{\Delta f S_n(f)}, \quad (15)$$

where S_n is the LISA power spectral density (PSD):

$$S_n(f) = \frac{20}{3} \frac{4S_{n,\text{acc}}(f) + S_{n,\text{sn}}(f) + S_{n,\text{omn}}(f)}{L^2} \times \left[1 + \left(\frac{f}{\frac{0.41c}{2L}} \right)^2 \right], \quad (16)$$

L corresponds to the detector arm length, and $S_{n,\text{acc}}$, $S_{n,\text{sn}}$, and $S_{n,\text{omn}}$ are the noise components due to low-frequency acceleration, shot

¹⁴The total LISA h_{rms} noise is the sum in quadrature of the instrumental rms noise and the confusion noise from unresolved galactic (Nelemans, Yungelson & Portegies Zwart 2001), extragalactic (Farmer & Phinney 2003), and white dwarf-white dwarf binaries. The number of these sources is expected to decrease as the LISA mission progresses and a larger number of foreground sources are detected and removed.

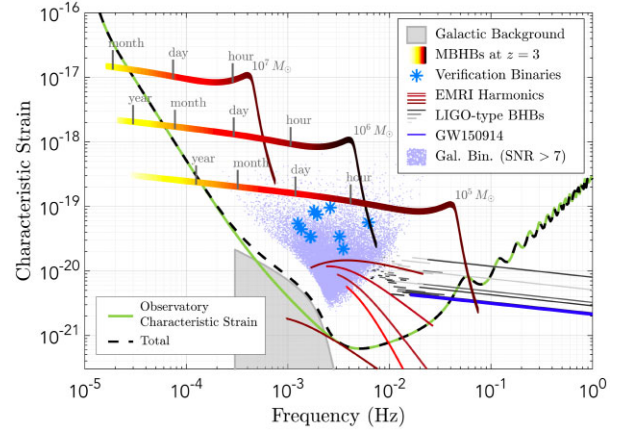


Figure 6. The yellow–red lines show the varying characteristic strains through the inspiral, merger, and ringdown phases for MBHB systems of total masses 10^5 , 10^6 , and $10^7 M_\odot$ from bottom to top. Expected sensitivity (green) with various possible sources in units of dimensionless characteristic strain amplitude for a three arm configuration of LISA. Plot taken from the LISA L3 mission proposal.¹⁵

noise and other measurement noise, respectively (Klein et al. 2016; Smith & Caldwell 2019), parametrized as¹⁶

$$S_{n,\text{acc}} = \frac{9 \times 10^{-30}}{(2\pi f)^4} \left(1 + \frac{10^{-4}}{f} \right) [\text{m}^2 \text{Hz}^{-1}],$$

$$S_{n,\text{sn}} = 2.22 \times 10^{-23} [\text{m}^2 \text{Hz}^{-1}],$$

$$S_{n,\text{omn}} = 2.65 \times 10^{-23} [\text{m}^2 \text{Hz}^{-1}]. \quad (17)$$

In an ideal experiment, to maximize the SNR, one should integrate equation (10) over the entire duration of the GW event. Most of the lifetime of a GW emitted by a MBHB is encompassed within the time interval between when the distance between the two MBHBs becomes close to the *hardening* radius r_h (the inspiral phase begins) and when it reaches the *innermost-stable circular orbit* radius r_{isco} (the merging phase begins). In this case, the integration limits should range between a minimum frequency $f_{\text{min}} = f_h$ at the r_h and the frequency $f_{\text{max}} = f_{\text{isco}}$ at the r_{isco} . However, in a real experiment, a GW event can be detected by LISA only if its frequency is included in the range $(10^{-4} - 1.0)$ Hz and the SNR overcomes a certain threshold (here taken as $\text{SNR}_{\text{thresh}} = 5$). We, thus consider as f_{min} the frequency at which $\text{SNR} > \text{SNR}_{\text{thresh}}$. We, hence calculate the SNR for each merger event in *AGN-fid*, and we find the results shown in the upper panels of Fig. 7.

The upper left-hand panel of Fig. 7 shows the trend of the SNR with redshift, without assuming any time delay. For a given chirp mass, the further the source is located, the lower is the SNR of the GW event. Furthermore, although h_c increases with the mass of the MBHB system, for a fixed redshift, the SNR is higher for sources with lower chirp mass. This trend occurs because low-mass binaries merge slower and enter in the LISA band sooner: hence, they stay in the LISA band for longer time and accumulate more SNR over their inspiraling lifetime. The maximum SNR that is resulting in this case is ~ 100 .

The upper right-hand panel of Fig. 7 shows instead the same trend of the SNR with redshift, when time delay due to dynamical friction

¹⁵https://www.elisascience.org/files/publications/LISA_L3_20170120.pdf

¹⁶These values hold for the current LISA design, that presents three spacecrafts connected by six links. Given the large uncertainties on the very-low frequency LISA sensitivity, we adopt a pessimistic cut at 10^{-4} Hz (Sesana et al. 2004).

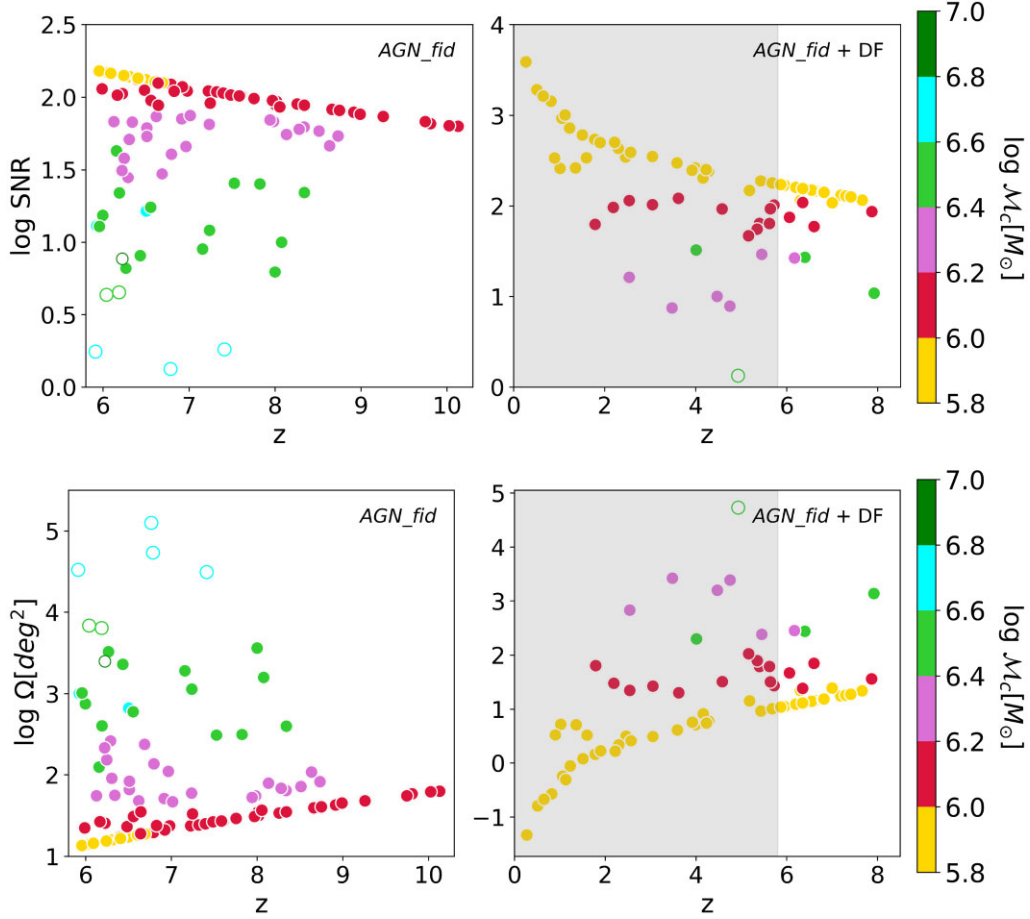


Figure 7. SNR ratio (upper panel) and angular resolution (lower panel) of MBHB mergers resulting from our simulations, with (without) delay times considered in right-hand (left) panel colour-coded according to their chirp mass. The shaded grey area represents the redshift range inaccessible to our simulations. Filled and empty circles represents detectable and undetectable events, respectively. The detectability threshold has been set to $\text{SNR}_{\text{thres}} = 5$.

is taken into account. We find that in this case events delayed at epochs $z < 6$ are characterized by SNRs that can be as high as $10^3 - 10^4$, while at $z > 6$ the highest SNR limit remains the same as *AGN_fid* without any delays ($\text{SNR} \sim 150$).

Finally, for the prospect of follow-up observations with electromagnetic telescopes, we calculate the LISA angular resolutions of LDEs to quantify how large is the region in the sky that must be covered by a telescope to detect the electromagnetic signals from merging MBHBs. We adopt the results found by McGee, Sesana & Vecchio (2020), which are derived from a range of population models: the median angular resolution Ω can be associated with the median SNR at which a merger is observed by the following relation:

$$\Omega \approx 0.5 \left(\frac{\text{SNR}}{10^3} \right)^{-7/4} \text{deg}^2. \quad (18)$$

The lower panel of Fig. 7 tracks the redshift evolution of the angular resolution for LDEs. Following the SNR trend, the optimal angular resolution is found for lower redshift (higher h_c) and smaller chirp mass (higher SNR) systems. This figure clearly shows that, although GW events from MBHB coalescence can be detected at high- z , their sky localization is poor (10 deg^2 in the most optimistic case; see also McWilliams et al. 2011), making follow-up observations in different EM bands challenging.

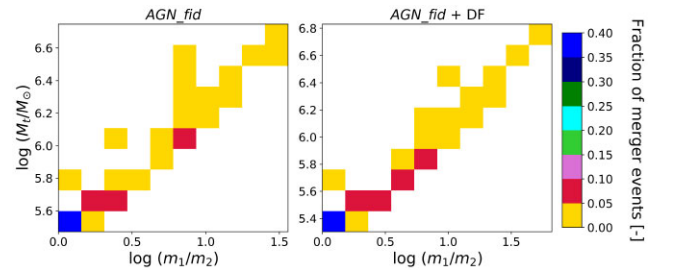


Figure 8. Fraction of detectable merger events as a function of the total mass of the system $M_t = m_1 + m_2$ and the mass ratio m_1/m_2 of the binary.

5.2 Mass ratio

Fig. 8 shows how the number of LDEs is distributed in terms of the ‘mass ratio’ (defined as m_1/m_2 , and shown in the x -axis) and the ‘total mass’ ($M_t = m_1 + m_2$, and shown in the y -axis), for *AGN_fid* with (right) and without (left) considering additional delay in the mergers due to dynamical friction.

The highest fraction of LDEs occurs for equi-mass binaries, in the low mass range ($M_t \lesssim 3 \times 10^6 M_\odot$), irrespective of consideration of delay time and is about 35 per cent–40 per cent of the total LDE population in each run. In other words, ‘just-seeded’ black holes in

Table 3. Detectability of MBHBs with LISA for the simulation runs adopted in this work (first column). The second column reports the total number of mergers in each run (N_{total}) while the third shows the LDEs fraction (f_{det}) in each run. The fourth (fifth, sixth, and seventh) column the chirp mass (observational time, observed frequency, and observed initial distance of separation) ranges for the LDEs. The eighth (ninth) column shows the fraction f_{out} (f_{und}) of MBHBs whose frequency is outside the LISA band (inside the LISA band but do not reach the required SNR threshold). In the last column, the minimum and maximum SNR accumulated by MBHBs, which merge within the LISA band but fail to reach the SNR threshold for detectability (SNR_{und}) are also shown.

Run	N_{total}	f_{det}	$\mathcal{M}_c[10^6 M_{\odot}]$	$t_{\text{obs}}[\text{days}]$	$\nu_{\text{obs}}[\text{mHz}]$	$R_{\text{obs}}[10^{-6}\text{pc}]$	$f_{\text{out}}[\text{mHz}]$	$f_{\text{und}}[\text{mHz}]$	SNR_{und}
<i>AGN_fid</i>	145	0.69	0.9–7.0	0.5–30	0.10–0.38	1.9–9.3	0.31	0.04	0.8–4.0
<i>AGN_fid + DF</i>	116	0.66	0.2–3.2	0.0–18	0.10–0.9	0.1–9.6	0.31	0.01	1.33
<i>BHs_noFB</i>	140	0.46	0.9–6.0	0.7–39	0.09–0.38	1.9–9.5	0.50	0.03	0.7–3.9
<i>AGNcone</i>	583	0.76	0.6–10	1.1–75	0.09–0.42	1.8–10.5	0.22	0.02	0.4–4.7
<i>AGNsphere</i>	415	0.78	0.6–7.5	0.6–54	0.09–0.40	1.6–8.4	0.22	0.005	1.6–3.7

binaries have a higher probability to coalesce as compared to black holes with higher mass, assembled by accretion and/or merging.

Furthermore, we note that for a fixed M_i , the larger is the mass ratio the higher is the fraction of mergers. This is simply due to the fact that the number of black holes decreases with increasing masses, thus for a fixed M_i , most of the mergers occurs for the BHs with lowest mass.

5.3 Time, frequency, and spatial scales of interest

Two MBHBs¹⁷ at a distance R require a certain time to merge through the emission of GWs. Such a time-scale is called *coalescing* time (t_{coal}) and defined as

$$t_{\text{coal}} = \frac{5}{256} \frac{c^5 R^4}{G^3 M_i^2 \mu}, \quad (19)$$

where R is the initial separation of the two merging MBHBs and $\mu = M_1 M_2 / M_i$ is the symmetric mass ratio. However, as already mentioned in Section 5.1, we can detect the GW signal only after $\text{SNR} > \text{SNR}_{\text{thresh}}$. We, thus compute the *observational* time as the time interval between the moment when $\text{SNR} > \text{SNR}_{\text{thresh}}$ and when the two MBHBs start merging:

$$t_{\text{obs}} = t_{\text{coal}} |_{\text{SNR} > \text{SNR}_{\text{thresh}}}, \quad (20)$$

which provides the period of time during which the GW event is actually observable. This time-scale provides the interval of time required to eventually trigger electromagnetic telescopes for follow-up observations to find the EM counterpart of the MBHB merging (see for example Loeb 2016), or to simply probe the host galaxies of the system.

To the observational time t_{obs} of a merger event we can associate the frequency ν_{obs} and the separation of the MBHBs such that $\text{SNR} > \text{SNR}_{\text{thresh}}$, which are related by the following expression:

$$\nu_{\text{obs}} = \pi^{-1} \left(\frac{GM_i}{R_{\text{obs}}^3} \right)^{1/2}. \quad (21)$$

We summarize the resulting scales of interest in Table 3.

6 CAVEATS AND DISCUSSION

The results presented in this work are affected by several limitations:

(i) Spatial resolution: the dynamical range required to study galaxy formation and BH co-evolution through numerical simulations is

¹⁷We only consider binary coalescences in our analysis, neglecting systems composed by three (or more) MBHBs.

extremely large, ranging from hundreds of Mpc scales (to search for massive DM haloes in N -body simulations) to sub-kpc scales (required to properly model gas accretion on to BHs and the subsequent feedback processes). This implies that the spatial resolution that is possible to achieve within a reasonable amount of computational time is typically limited to tens of pc scale,¹⁸ namely more than 7 orders of magnitude larger than the typical scales of interest for GW studies (see Table 3).

(ii) Temporal resolution: a discussion similar to the spatial resolution can be done for what concerns the temporal resolution. Black hole properties in our simulations are evolved with time-steps of ~ 0.01 –1 Myr, which are several orders of magnitude larger than the typical scales of interest for GW studies (see Table 3). The situation gets worse if we consider the time interval between two snapshots (\sim tens Myr). This implies that we are not associating an LDE to the galaxy properties at the time of the coalescence (see Section 4.1).

(iii) Seeding prescription: our simulations are based on a seeding prescription that mimicks the DCBH, given the mass of the seeds that we choose. However, the formation of a DCBH is governed by a complex network of physical processes (i.e. H_2 formation, metal enrichment, and radiative transfer) that is impossible to take into account self-consistently in cosmological zoom-in HDS. Our seeding prescription (namely a BH of mass $10^5 M_{\odot}$ in each $10^9 M_{\odot}$ DM halo) is, thus certainly overestimating the number of MBHBs in B18 simulations (see e.g. Vito et al. 2022) and consequently the expected number of merger events reported here.

For the reasons reported above, the merger rate computed in this work represents a solid *upper limit* on the number of merger events that LISA will be able to detect.

7 SUMMARY AND CONCLUSIONS

In this work, we adopted a suite of cosmological zoom-in HDS of galaxy formation and BH co-evolution, developed with the GADGET-3 code, and characterized by different numerical resolutions and star formation/AGN feedback prescriptions. Our simulations are based on a seeding prescription such that $M_{\text{seed}} \sim 10^5 M_{\odot}$ BHs are planted in $M_{\text{DM}} \sim 10^9 M_{\odot}$ DM haloes; furthermore, if two BHs are at a distance smaller than the smoothing length of the simulations and their relative velocity is smaller than the local sound speed, they are assumed to merge instantaneously. We use these simulations to investigate the coalescence of MBHB

¹⁸The largest spatial resolution achieved so far by zoom-in cosmological simulations is ~ 15 pc of a $z = 7$ quasar (Lupi et al. 2019).

($M_{\text{BH}} \gtrsim 10^6 M_{\odot}$) at $6 < z < 10$ and to compute their GW properties. We summarize below the findings of our work and draw the main conclusions arising from these results.

(i) Merger rate: we calculated the merger rates of MBHBs with different AGN feedback scenarios and numerical resolution: thermal (*AGN_fid* and *BHs.noFB* by V21, $m_{\text{DM}} = 1.5 \times 10^6 M_{\odot}$) and kinetic (*AGNcone* and *AGNsphere* by B18, $m_{\text{DM}} = 7.5 \times 10^6 M_{\odot}$). We found that the merger rate strongly depends on the numerical resolution adopted, ranging within 80–300 events per year in the V21 and B18, respectively at redshift 6. Furthermore, the merger rates of MBHBs at fixed resolution depend on the feedback recipe implemented: the *AGN_fid* model predicts a merger rate that is a factor $\times 2$ higher than *BHs.noFB* at $z \sim 6$ for chirp masses in the range $10^6 < \mathcal{M}_c < 10^7 M_{\odot}$; analogously, the *AGNcone* model predicts a merger rate that is a factor $\times 3$ higher than *AGNsphere* at $z \sim 6$ both for low ($\mathcal{M}_c < 10^6 M_{\odot}$) and high ($\mathcal{M}_c > 10^7 M_{\odot}$) chirp masses. Different feedback prescriptions and numerical resolutions also affect the epoch of the furthest GW signal detectable with LISA: in the V21 simulations, for $\mathcal{M}_c > 10^7 M_{\odot}$, the furthest GW signal occurs at $z \sim 7.7$ – 8.0 while in the B18 at $z \sim 7.3$ – 7.5 . We discussed in details all the physical and numerical explanations for these trends and we underlined the several motivations that make our predictions stringent upper limits to the actual merger rates that are expected to be observed.

(ii) Halo bias: our merger rate predictions are biased since they are based on zoom-in simulations targeting massive DM haloes. To quantify this bias, we adopted the SAM by Barausse (2012) and showed that a MBH model like ours (with heavy BH seeds and instantaneous mergers) overpredicts the actual merger rate by a factor of ~ 20 at comparable redshifts ($z > 6$). For lighter seeds as well as heavy seed models with time delays included, the merger rate overprediction still remains at a factor between 20 and 30.

(iii) Delays in MBHB mergers: we corrected in post-processing the coalescing time of MBHB mergers including a time delay due to dynamical friction from the surrounding stars. We found that, if this delay is considered, 83 per cent of MBHBs will merge within the Hubble time, but only 21 per cent of them, will merge within 1 Gyr, namely the age of the Universe at $z > 6$.

(iv) SNR ratio and angular resolution: taking into account the LISA frequency bands, we calculated the SNR ratio and the angular resolution of the GW events predicted by our fiducial run, *AGN_fid*. The fraction of LISA detectable events with high SNR ratio ($\text{SNR} > 5$) ranges between 66 per cent and 69 per cent depending on the inclusion of time delays in post-processing. The largest SNR is reached in the case of low chirp masses ($< 10^6 M_{\odot}$) which, although being characterized by a smaller characteristic strain, remain in the LISA band for a longer time, thus decreasing the noise and increasing the SNR. These systems are however, characterized by very low angular resolutions (10 deg^2).

(v) Mass ratio of LDEs: we computed the distribution of the mass ratio of the LDEs in the *AGN_fid* run and we found that the maximum number of mergers occurs for equi-massed binaries, which are ‘just-seeded’ (i.e. with $M_{\text{BH}} \sim 10^5 M_{\odot}$). For a fixed total mass of a MBHB, the number of mergers increases with increasing mass ratio, which can be attributed to hierarchical structure formation: several low-mass MBHBs merge (and accrete) to form more massive MBHBs. This remained true even when further time delays in merging are considered.

One of the main goals of our study was to quantify the range of uncertainties on the merger rate of LISA detectable events. We find that considerable different merger rates result from the simulations

in our suite. To get a more reliable constraint on this important issue, the shortage of currently adopted models should be addressed. For what concerns the EM signals arising from LDEs, the main challenge remains the poor LISA sky localization that in the most optimistic case ($\mathcal{M}_c \sim 10^6 M_{\odot}$ at $z = 6$) is around 10 deg^2 . For this reason, it is important to further investigate the EM properties of LDEs to search for eventual, unique signatures from MBH coalescences to maximize the chances of their detections, which we address in our future work.

ACKNOWLEDGEMENTS

The authors thank Andrea Pallottini, Hamsa Padmanabhan, and Alessandro Lupi for their helpful insights and comments. SG acknowledges support from the ASI-INAF (Agenzia Spaziale Italiana-Istituto Nazionale di Astrofisica) grant (n. 2018-31-HH.0) and PRIN-MIUR (Progetti di rilevante interesse nazionale - Ministero dell’Istruzione e del Merito) grant (2017PH3WAT_005). MV is supported by the Alexander von Humboldt Stiftung and the Carl Friedrich von Siemens Stiftung. MV also acknowledges support from the Excellence Cluster ORIGINS, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy (EXC-2094 - 390783311). PB acknowledges support from the Brazilian funding agency FAPESP (Fundaç o de Amparo   Pesquisa de Estado de S o Paulo, grants 2016/01355-5, 2016/22183-8). D.I.V. acknowledges the financial support provided under the European Union’s H2020 ERC (European Research Council) Consolidator Grant ‘Binary Massive Black Hole Astrophysics’ (B Massive, Grant Agreement: 818691).

DATA AVAILABILITY STATEMENT

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

REFERENCES

- Amaro-Seoane P. et al., 2022, *Liv. Rev. Relat.*, 26, 2
 Arun K. G. et al., 2009, *Class. Quant. Grav.*, 26, 094027
 Arzoumanian Z. et al., 2018, *ApJ*, 859, 47
 Barai P., de Gouveia Dal Pino E. M., 2019, *MNRAS*, 487, 5549
 Barai P., Gallerani S., Pallottini A., Ferrara A., Marconi A., Ciccone C., Maiolino R., Carniani S., 2018, *MNRAS*, 473, 4003
 Barausse E., 2012, *MNRAS*, 423, 2533
 Barausse E., Dvorkin I., Tremmel M., Volonteri M., Bonetti M., 2020, *ApJ*, 904, 16
 Barkana R., Loeb A., 2001, *Phys. Rep.*, 349, 125
 Begelman M. C., Volonteri M., Rees M. J., 2006, *MNRAS*, 370, 289
 Bhowmick A. K. et al., 2022, *MNRAS*, 516, 138
 Binney J., Tremaine S., 2008, *Galactic Dynamics*, 2nd edn. Princeton Univ. Press, Princeton, NJ
 Blanchet L., Damour T., Iyer B. R., Will C. M., Wiseman A. G., 1995, *Phys. Rev. Lett.*, 74, 3515
 Blecha L. et al., 2015, *MNRAS*, 456, 961
 Blumenthal G. R., Faber S., Primack J. R., Rees M. J., 1984, *Nature*, 311, 517
 Bondi H., 1952, *MNRAS*, 112, 195
 Bondi H., Hoyle F., 1944, *MNRAS*, 104, 273
 Bonetti M., Sesana A., Haardt F., Barausse E., Colpi M., 2019, *MNRAS*, 486, 4044
 Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
 Chabrier G., 2003, *PASP*, 115, 763
 Chandrasekhar S., 1943, *ApJ*, 97, 255
 Ciardi B., Loeb A., 2000, *ApJ*, 540, 687

- eLISA Consortium et al., 2013, preprint ([arXiv:1305.5720](https://arxiv.org/abs/1305.5720))
- Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015a, *MNRAS*, 450, 1514
- Correa C. A., Wyithe J. S. B., Schaye J., Duffy A. R., 2015b, *MNRAS*, 450, 1521
- Cutler C., Flanagan É. E., 1994, *Phys. Rev. D*, 49, 2658
- Davies M. B., Miller M. C., Bellovary J. M., 2011, *ApJ*, 740, L42
- Dayal P., Rossi E. M., Shiralilou B., Piana O., Choudhury T. R., Volonteri M., 2019, *MNRAS*, 486, 2336
- DeGraf C., Sijacki D., 2019, *MNRAS*, 491, 4973
- DeGraf C., Sijacki D., Di Matteo T., Holley-Bockelmann K., Snyder G., Springel V., 2021, *MNRAS*, 503, 3629
- Devecchi B., Volonteri M., Rossi E. M., Colpi M., Portegies Zwart S., 2012, *MNRAS*, 421, 1465
- Di Matteo T., Khandai N., DeGraf C., Feng Y., Croft R. A. C., Lopez J., Springel V., 2012, *ApJ*, 745, L29
- Dong-Páez C. A., Volonteri M., Beckmann R. S., Dubois Y., Trebitsch M., Mangiagli A., Vergani S. D., Webb N. A., 2023, *A&A*, 673, A120
- Edgar R., 2004, *New Astron. Rev.*, 48, 843
- Eisenstein D. J., Loeb A., 1995, *ApJ*, 443, 11
- Fan X. et al., 2006, *ApJ*, 132, 117
- Farmer A. J., Phinney E. S., 2003, *MNRAS*, 346, 1197
- Ferrara A., Salvadori S., Yue B., Schleicher D., 2014, *MNRAS*, 443, 2410
- Flanagan É. É., Hughes S. A., 1998, *Phys. Rev. D*, 57, 4535
- Haehnelt M. G., 1994, *MNRAS*, 269, 199
- Hahn O., Abel T., 2011, *MNRAS*, 415, 2101
- Hartwig T., Agarwal B., Regan J. A., 2018, *MNRAS*, 479, L23
- Hawking S. W., Israel W., 1989, *Three Hundred Years of Gravitation*. Cambridge Univ. Press, Cambridge
- Heger A., Fryer C. L., Woosley S. E., Langer N., Hartmann D. H., 2003, *ApJ*, 591, 288
- Hirano S., Hosokawa T., Yoshida N., Omukai K., Yorke H. W., 2015, *MNRAS*, 448, 568
- Hoyle F., Lyttleton R. A., 1939, *Proc. Camb. Philos. Soc.*, 35, 405
- Izquierdo-Villalba D., Sesana A., Bonoli S., Colpi M., 2021, *MNRAS*, 509, 3488
- Jaffe A. H., Backer D. C., 2003, *ApJ*, 583, 616
- Jiang C. Y., Jing Y. P., Lin W. P., 2010, *A&A*, 510, A60
- Johnson J. L., Whalen D. J., Fryer C. L., Li H., 2012, *ApJ*, 750, 66
- Katz M. L., Larson S. L., 2019, *MNRAS*, 483, 3108
- Katz M. L., Kelley L. Z., Dospoulou F., Berry S., Blecha L., Larson S. L., 2019, *MNRAS*, 491, 2301
- Kelley L. Z., Blecha L., Hernquist L., 2016, *MNRAS*, 464, 3131
- Klein A. et al., 2016, *Phys. Rev. D*, 93, 092004
- Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
- Kormendy J., Gebhardt K., 2001, *AIP Conf. Proc. Vol. 586, Relativistic Astrophysics: 20th Texas Symposium*. Am. Inst. Phys., New York, p. 363
- Koushiappas S. M., Bullock J. S., Dekel A., 2004, *MNRAS*, 354, 292
- Krolik J. H., Volonteri M., Dubois Y., Devriendt J., 2019, *ApJ*, 879, 110
- Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J., 2013, *MNRAS*, 433, 1607
- Lentati L. et al., 2015, *MNRAS*, 453, 2577
- Lodato G., Natarajan P., 2006, *MNRAS*, 371, 1813
- Loeb A., 2016, *ApJ*, 819, L21
- Loeb A., Rasio F. A., 1994, *ApJ*, 432, 52
- Lupi A., Colpi M., Devecchi B., Galanti G., Volonteri M., 2014, *MNRAS*, 442, 3616
- Lupi A., Volonteri M., Decarli R., Bovino S., Silk J., Bergeron J., 2019, *MNRAS*, 488, 4004
- Madau P., Pozzetti L., Dickinson M., 1998, *ApJ*, 498, 106
- Magorrian J. et al., 1998, *ApJ*, 115, 2285
- Mapelli M., 2016, *MNRAS*, 459, 3432
- McGee S., Sesana A., Vecchio A., 2020, *Nature Astron.*, 4, 26
- McWilliams S. T., Lang R. N., Baker J. G., Thorpe J. I., 2011, *Phys. Rev. D*, 84, 064003
- Mikkola S., Valtonen M. J., 1992, *MNRAS*, 259, 115
- Murante G., Monaco P., Giovalli M., Borgani S., Diaferio A., 2010, *MNRAS*, 405, 1491
- Murante G., Monaco P., Borgani S., Tornatore L., Dolag K., Goz D., 2015, *MNRAS*, 447, 178
- Navarro J. F., Frenk C. S., White S. D. M., 1996, *ApJ*, 462, 563
- Nelemans G., Yungelson L. R., Portegies Zwart S. F., 2001, *A&A*, 375, 890
- Nelson D. et al., 2015, *Astron. Comput.*, 13, 12
- Ostriker E. C., 1999, *ApJ*, 513, 252
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
- Pillepich A. et al., 2018, *MNRAS*, 473, 4077
- Planck Collaboration, 2016, *A&A*, 594, A13
- Press W. H., Schechter P., 1974, *ApJ*, 187, 425
- Quinlan G. D., 1996, *New Astron.*, 1, 35
- Reines A. E., Greene J. E., Geha M., 2013, *ApJ*, 775, 116
- Reinoso B., Schleicher D. R. G., Fellhauer M., Klessen R. S., Boekholt T. C. N., 2018, *A&A*, 614, A14
- Schaye J. et al., 2015, *MNRAS*, 446, 521
- Salcido J., Bower R. G., Theuns T., McAlpine S., Schaller M., Crain R. A., Schaye J., Regan J., 2016, *MNRAS*, 463, 870
- 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., Volonteri M., 2005, *ApJ*, 623, 23
- Sesana A., Volonteri M., Haardt F., 2007, *MNRAS*, 377, 1711
- Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192
- Sesana A., Shankar F., Bernardi M., Sheth R. K., 2016, *MNRAS*, 463, L6
- Shang C., Bryan G. L., Haiman Z., 2010, *MNRAS*, 402, 1249
- Shannon R. M. et al., 2015, *Science*, 349, 1522
- Sijacki D., Springel V., Haehnelt M. G., 2009, *MNRAS*, 400, 100
- Silk J., 2017, *ApJ*, 839, L13
- Silk J., Rees M. J., 1998, *A&A*, 331, L1
- Śądowski A., Gaspari M., 2017, *MNRAS*, 468, 1398
- Smith T. L., Caldwell R. R., 2019, *Phys. Rev. D*, 100, 104055
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Springel V., Di Matteo T., Hernquist L., 2005, *MNRAS*, 361, 776
- Steinborn L. K., Dolag K., Hirschmann M., Prieto M. A., Remus R.-S., 2015, *MNRAS*, 448, 1504
- Tanaka T., Haiman Z., 2009, *ApJ*, 696, 1798
- Tornatore L., Borgani S., Dolag K., Matteucci F., 2007, *MNRAS*, 382, 1050
- Valentini M., Murante G., Borgani S., Monaco P., Bressan A., Beck A. M., 2017, *MNRAS*, 470, 3167
- Valentini M., Borgani S., Bressan A., Murante G., Tornatore L., Monaco P., 2019, *MNRAS*, 485, 1384
- Valentini M. et al., 2020, *MNRAS*, 491, 2779
- Valentini M., Gallerani S., Ferrara A., 2021, *MNRAS*, 507, 1
- Valiante R. et al., 2020, *MNRAS*, 500, 4095
- Vito F., Di Mascia F., Gallerani S., Zana T., Ferrara A., Carniani S., Gilli R., 2022, *MNRAS*, 514, 1672
- Vogelsberger M. et al., 2014, *Nature*, 509, 177
- Volonteri M., Bellovary J., 2012, *Rep. Progr. Phys.*, 75, 124901
- Volonteri M., Haardt F., Madau P., 2003, *ApJ*, 582, 559
- Volonteri M., Reines A. E., Atek H., Stark D. P., Trebitsch M., 2017, *ApJ*, 849, 155
- Volonteri M. et al., 2020, *MNRAS*, 498, 2219
- Weinberger R. et al., 2017, *MNRAS*, 465, 3291
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wiersma R. P. C., Schaye J., Smith B. D., 2009, *MNRAS*, 393, 99
- Woods T. E. et al., 2019, *PASA*, 36, e027
- Wyithe J. S. B., Loeb A., 2003, *ApJ*, 590, 691
- Yoshida N., Omukai K., Hernquist L., 2008, *Science*, 321, 669
- Yue B., Ferrara A., Salvaterra R., Xu Y., Chen X., 2013, *MNRAS*, 433, 1556
- Zana T., Gallerani S., Carniani S., Vito F., Ferrara A., Lupi A., Di Mascia F., Barai P., 2022, *MNRAS*, 513, 2118

APPENDIX A: CONSTRAINTS FROM PTA

We compare our predictions with upper limits placed by several Pulsar Timing Array (PTA) observations. PTAs are sensitive to GWs

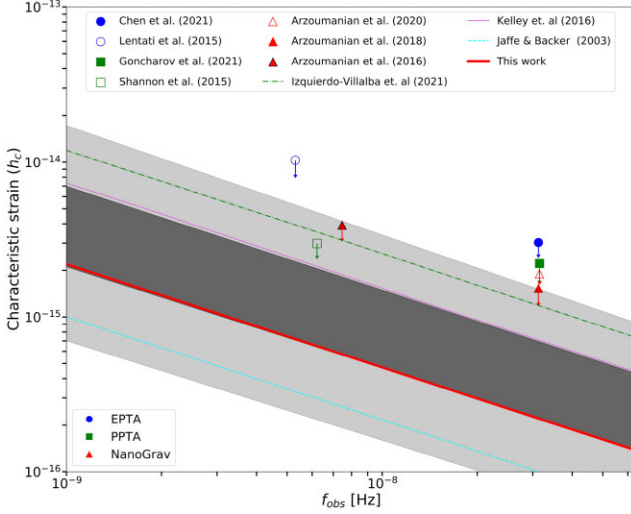


Figure A1. Stochastic GWB spectrum calculated from our fiducial model (red solid line). The circles, squares, and triangles represent EPTA, PPTA, and NanoGrav data, respectively. Dark (light) grey shaded region depicts the 1σ (2σ) confidence level of the predictions by Sesana et al. (2016). The green dot-dashed line shows the results by Izquierdo-Villalba et al. (2021): $A_{\text{yr}^{-1}} \sim 1.2 \times 10^{-15}$. The solid pink line refers to the Illustris predictions of $A_{\text{yr}^{-1}} \sim 7.1 \times 10^{-16}$ (Kelley et al. 2016). The cyan dashed line shows the results by Jaffe & Backer (2003): $A_{\text{yr}^{-1}} \sim 1 \times 10^{-16}$.

with frequencies between 10^{-9} and 10^{-7} Hz. In this nano-Hz regime, the signal mostly arises from stochastic GWB produced by the incoherent superposition of GWs from the population of inspiralling MBHBs overlapping in frequencies. The characteristic strain arising from this stochastic GWB can be written as (Sesana, Vecchio & Colacino 2008)

$$h_c^2(f) = \frac{4G^{5/3}}{f^2 c^2 \pi} \iint \frac{dz d\mathcal{M}}{(1+z)} \frac{d^2 N}{dz d\mathcal{M}} \frac{dE_{\text{GW}}(\mathcal{M})}{d \ln f_r}, \quad (\text{A1})$$

where $d^2 N/dz d\mathcal{M}$ is the comoving number density of MBHB merger per unit redshift, and rest-frame chirp mass, f is the frequency of the GWs in the observer frame, and $dE_{\text{GW}}/d \ln f_r$ is the energy emitted per logarithmic rest-frame frequency, f_r .

Assuming the inspiralling population of MBHBs in the PTA band are in perfect circular orbits, equation (A1) can be re-written as

$$h_c^2(f) = \frac{4G^{5/3} f^{-4/3}}{3c^2 \pi^{1/3}} \iint dz d\mathcal{M} \frac{d^2 N}{dz d\mathcal{M}} \frac{\mathcal{M}^{5/3}}{(1+z)^{1/3}}. \quad (\text{A2})$$

This type of relation is typically written as

$$h_c(f) = A \left(\frac{f}{f_0} \right)^{-2/3}, \quad (\text{A3})$$

where A is the amplitude of the signal at the reference frequency f_0 , which is usually normalized at $f_0 = 1 \text{ yr}^{-1}$, and $A(f_0 = 1 \text{ yr}^{-1})$ is typically denoted as $A_{\text{yr}^{-1}}$.

We present our predictions in Fig. A1, where we apply a bias correction factor of 20 (see Section 3.2). We find $A_{\text{yr}^{-1}} \sim 2 \times 10^{-16}$, which is below the upper limits placed by the NANOGrav (Arzoumanian et al. 2018), EPTA (Lentati et al. 2015) observations, and the PPTA observations (Shannon et al. 2015). We emphasize that our predictions do not include any contribution from sources at $z < 6$, which are instead expected to dominate the background (e.g. Izquierdo-Villalba et al. 2021).

APPENDIX B: TIME DELAY DUE TO STELLAR HARDENING

Stellar hardening is the process by which stars interact with the MBHB and gradually become more tightly bound to it over time. For already bound MBHBs, the gravitational field can be strong enough to disrupt the orbits of nearby stars. As a result, some of these stars can be captured by the MBHB and start to orbit around it (Mikkola & Valtonen 1992). As these stars continue to interact with the binary, they can extract energy and angular momentum from the MBHB, causing it to become more tightly bound (Quinlan 1996).

We compute the stellar hardening time-scale following Sesana & Khan (2015):

$$t_{\text{stellar}} = 15.18 \text{ Gyr} \left(\frac{\sigma_{\text{inf}}}{\text{km s}^{-1}} \right) \left(\frac{\rho_{\text{inf}}}{\text{M}_{\odot} \text{pc}^{-3}} \right)^{-1} \left(\frac{a_{\text{gw}}}{10^{-3} \text{ pc}} \right)^{-1}, \quad (\text{B1})$$

where σ_{inf} and ρ_{inf} are the velocity dispersion and the stellar density at the sphere of influence,¹⁹ and a_{gw} is the transition separation of stellar hardening and GW hardening at which the binary spends most of its time:

$$\sigma_{\text{inf}} = (GM_t/r_{\text{inf}})^{1/2}, \quad (\text{B2})$$

$$\rho_{\text{inf}} = \frac{(3-\gamma)M_* r_{\text{inf}}^{-\gamma}}{8\pi R_{\text{eff}}^{3-\gamma}}, \quad (\text{B3})$$

$$a_{\text{gw}} = 2.64 \times 10^{-2} \text{ pc} \times \left[\frac{\sigma_{\text{inf}}}{\text{km s}^{-1}} \frac{\text{M}_{\odot} \text{pc}^{-3}}{\rho_{\text{inf}}} \frac{15}{H} \left(\frac{M_{\text{BHp}} M_{\text{BHs}} M_t}{2 \times 10^{24} \text{ M}_{\odot}^3} \right) \right]^{1/5}, \quad (\text{B4})$$

where r_{inf} is the radius containing twice the binary mass in stars:

$$r_{\text{inf}} = R_{\text{eff}} \left(\frac{4M_t}{M_*} \right)^{1/(3-\gamma)}, \quad (\text{B5})$$

and we assume the index $\gamma = 2$ (Volonteri et al. 2020) and $H = 15$ (Sesana & Khan 2015).

Finally, the total time of the MBHB merger including dynamical friction and stellar hardening is given by:

$$t_{\text{tot,stellar}} = t_{\text{tot,df}} + t_{\text{stellar}} \quad (\text{B6})$$

According to the formalism reported above, we find that t_{stellar} varies in the range 1–10 Gyr, since central stellar densities are in the range $40\text{--}600 \text{ M}_{\odot} \text{pc}^{-3}$. These ρ_{inf} values are not a fair representation, likely because the resolution of our simulations do not allow us to properly determine the matter distribution on such small scales. For comparison, in Sgr A* $\rho_{\text{inf}} \sim 7 \times 10^4 \text{ M}_{\odot} \text{pc}^{-3}$. In addition, galaxies at high redshift are likely more centrally concentrated than local galaxies. Thus, we would expect ρ_{inf} values even larger than what is found in Sgr A*. For all these reasons, we do not include the delay due to stellar hardening in our post-processing calculations.

APPENDIX C: GW DETECTABLES FOR DIFFERENT SIMULATION RUNS

Here, we show the comparison of different GW detectables for the simulation runs *AGNcone*, *AGNsphere*, and *BHs.noFB* compared with *AGNfid*. In all cases we report the results for a biased halo with no delays in post-processing. Fig. C1 shows the redshift distribution

¹⁹The sphere of influence is approximated as the sphere containing twice the binary mass in stars.

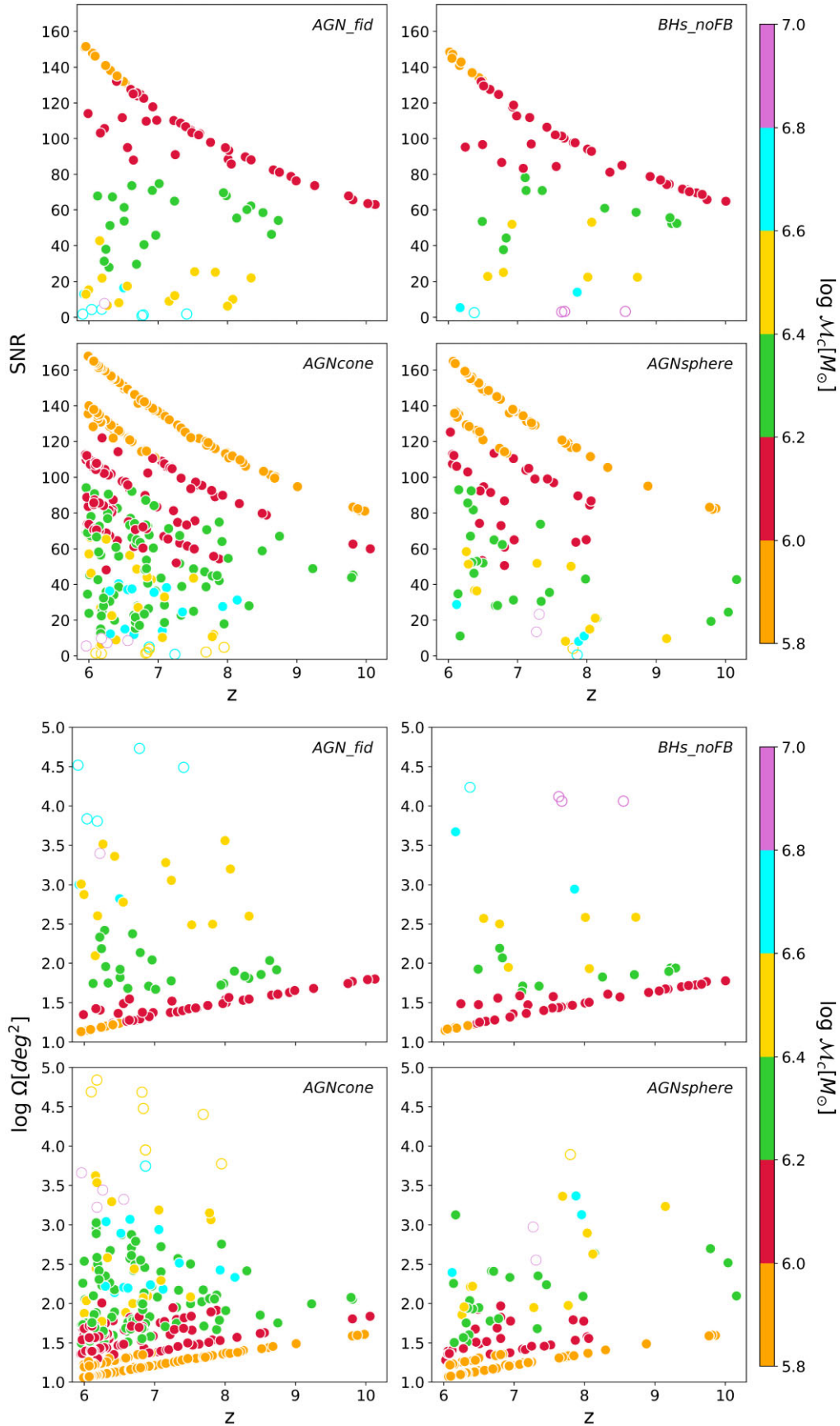


Figure C1. SNR ratio (upper panel) and angular resolution (lower panel) of MBHB mergers resulting from different runs of our simulations, colour-coded according to their chirp mass. Filled and empty circles represents detectable and undetectable events, respectively. The detectability threshold has been set to $\text{SNR}_{\text{thres}} = 5$.

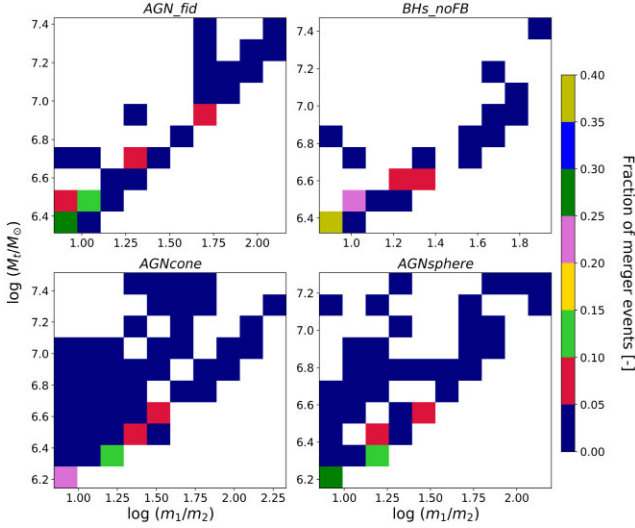


Figure C2. Fraction of detectable merger events as a function of the total mass of the system $M_t = m_1 + m_2$ and the mass ratio m_1/m_2 of the binary.

of SNR (upper panel) and angular resolution (lower panel) of the GW events for different chirp masses in our four different simulation suits. The overall trend in the SNR and resolution distribution is explained in Section 5.1. Since only in the B18 simulations

the chirp mass PDF is populated for $\mathcal{M}_c < 10^6 M_\odot$, this explains why for these simulations the SNR predicted reaches the highest value ($\text{SNR} > 160$). Apart from this aspect (and the different number of events predicted, already discussed in Section 3.1), we do not find in the SNR predictions huge differences among different models. Consequently, the corresponding angular resolutions of event the loudest GW events remain quite poor ($\sim 10 \text{ deg}^2$) for all models.

In Fig. C2, we also show the fraction of mergers are distributed in the ‘mass ratio’ and ‘total mass’ plane, in the four simulation runs analysed in this work. The different coverage of the ‘mass ratio’–‘total mass’ plane simply reflects the different numbers of LDEs in the simulations runs: as shown in Table 3 a higher number of LDEs occurs in the *AGNcone* run, which shows the most densely populated ‘mass ratio’–‘total mass’ plane, while the smaller number of LDEs is predicted in the *BHs_noFB* run.

As seen in Section 5.2, highest fraction of LDEs occurs for equi-mass binaries, in the low mass range ($M_t \lesssim 3 \times 10^6 M_\odot$), independent of the feedback implemented and the simulation resolution and it ranges between 20 percent and 40 percent of the total LDE population in each run.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.