

Premature supermassive black hole mergers in cosmological simulations of structure formation

Stephanie Buttigieg^{1,2,★}, Debora Sijacki^{1,2}, Christopher J. Moore^{1,2,3} and Martin A. Bourne^{1,2,4}

¹*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

²*Kavli Institute for Cosmology (KICC), University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

³*Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*

⁴*Centre for Astrophysics Research, Department of Physics, Astronomy and Mathematics, University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK*

Accepted 2025 August 8. Received 2025 July 21; in original form 2025 April 24

ABSTRACT

The co-evolution of massive black holes (BHs) and their host galaxies is well established within the hierarchical galaxy formation paradigm. Large-scale cosmological simulations are an ideal tool to study the repeated BH mergers, accretion and feedback that conspire to regulate this process. While such simulations are of fundamental importance for understanding the complex and intertwined relationship between BHs and their hosts, they are plagued with numerical inaccuracies at the scale of individual BH orbits. To quantify this issue, taking advantage of the $(100 h^{-1} \text{ cMpc})^3$ FABLE simulation box, we track all individual BH mergers and the corresponding host galaxy mergers as a function of cosmic time. We demonstrate that BH mergers frequently occur prematurely, well before the corresponding merger of the host galaxies is complete, and that BHs are sometimes erroneously displaced from their hosts during close galaxy encounters. Correcting for these artefacts results in substantial macrophysical delays, spanning over several Gyrs, which are additional to any microphysical delays arising from unresolved BH binary hardening processes. We find that once the macrophysical delays are accounted for, high-mass BH merger events are suppressed, affecting the predictions for the BH population that may be observable with *LISA* (*Laser Interferometer Space Antenna*) and pulsar timing arrays. Furthermore, including these macrophysical delays leads to an increase in the number of observable dual active galactic nuclei, especially at lower redshifts, with respect to FABLE. Our results highlight the pressing need for more accurate modelling of BH dynamics in cosmological simulations of galaxy formation as we prepare for the multimessenger era.

Key words: gravitational waves – methods: numerical – galaxies: kinematics and dynamics – quasars: supermassive black hole – black hole mergers.

1 INTRODUCTION

Observations provide compelling evidence that most, if not all, massive galaxies host a central supermassive black hole (SMBH) with masses ranging from 10^6 to $10^{11} M_\odot$ (Magorrian et al. 1998; Richstone et al. 1998; Kormendy & Ho 2013). Furthermore, these observations suggest a relation between SMBHs and their host galaxies, with the SMBH mass being found to correlate with key stellar properties, such as bulge luminosity, mass, and velocity dispersion (Magorrian et al. 1998; Gebhardt et al. 2000; Ferrarese 2002; Tremaine et al. 2002; Marconi & Hunt 2003; Häring & Rix 2004; Gültekin et al. 2009; Kormendy & Ho 2013; McConnell & Ma 2013). These relations suggest an evolutionary link between SMBHs and their host galaxies, influenced by processes such as galaxy mergers, active galactic nuclei (AGNs) and stellar feedback (Di Matteo, Springel & Hernquist 2005; Sijacki et al. 2007; Hopkins et al. 2008; Somerville et al. 2008; Dubois et al. 2016; Terrazas et al. 2020). However, significant uncertainties remain in these scaling relations

and their evolution across cosmic time (Merloni et al. 2009), with observations from the *James Webb Space Telescope* (*JWST*) starting to provide better constraints on these relations at higher redshifts (see e.g. Maiolino et al. 2024).

In the Λ -cold dark matter model, galaxies are expected to undergo mergers throughout the history of the Universe. When both galaxies involved in a merger host a SMBH, it is expected that a SMBH binary will form, which may eventually lead to coalescence. Before the two merging galaxies coalesce into a single system, it is the galaxy merger process itself that drives the shrinking of the SMBH separation, as the black holes (BHs) remain gravitationally bound to the centres of their respective host galaxies. Once the galaxies merge to form a unified remnant, the continued orbital decay of the SMBH binary is governed by a series of physical processes operating across different time and length scales, as outlined in the seminal work of Begelman, Blandford & Rees (1980).

During the galaxy merger, the two BHs are still an unbound pair and migrate toward the centre of the host galaxy via dynamical friction (DF, Chandrasekhar 1943; Binney & Tremaine 2011). As a BH moves through a background of gas, stars, and dark matter (DM), it generates a wake that exerts a gravitational drag force on the BH,

* E-mail: sb2583@cam.ac.uk

causing it to slow down (Chandrasekhar 1943; Antonini & Merritt 2012; Dosopoulou & Antonini 2017). When the binding energy of the BH pair exceeds the average kinetic energy of the surrounding stars, the system transitions into a ‘hard’ binary (Binney & Tremaine 2011). At this stage, DF becomes inefficient, and one of the primary mechanisms for orbital energy extraction is stellar scattering. Stars with sufficiently low angular momentum, residing in the so-called loss-cone of parameter space, interact with the binary. These interactions transfer energy from the binary to the stars, ejecting them at high velocities and shrinking the binary orbit in the process (Quinlan 1996; Khan et al. 2013). Additionally, in gas-rich environments, a circumbinary disc (CBD) can exert gravitational torques on the BHs, facilitating orbital decay at separations of approximately 0.1–10 pc (Pringle 1991; Artymowicz & Lubow 1994; Gould & Rix 2000; Armitage & Natarajan 2005; Lodato et al. 2009), with gas residing within the ‘cavity’ of the CBD, such as streams and mini-discs, being shown to have significant impact in this regime (Franchini, Lupi & Sesana 2022; Siwek, Weinberger & Hernquist 2023; Bourne et al. 2024). At sufficiently small separations ($\lesssim 10^{-3}$ pc), gravitational wave (GW) emission becomes the dominant mechanism for energy and angular momentum loss, driving the binary to coalescence (Peters 1964). Throughout these hardening mechanisms, the BHs may continue to accrete material, leading to changes in their mass and spin as they evolve toward their final pre-merger states (Bardeen & Petterson 1975; King & Kolb 1999; Merritt & Milosavljević 2005; Gerosa et al. 2015; Siwek, Kelley & Hernquist 2020; Bourne et al. 2024).

JWST is revolutionizing our understanding of the high-redshift SMBH population, enabling detailed studies of the redshift evolution of SMBH–host galaxy relationships out to redshifts as high as $z \approx 10$ (Pacucci et al. 2023; Bogdan et al. 2024; Juodžbalis et al. 2024; Maiolino et al. 2024), and perhaps even shedding light on the seeding mechanisms of SMBHs. Notably, *JWST* has identified dual AGN at rates an order of magnitude higher than predictions from cosmological simulations (Perna et al. 2023, 2025; Übler et al. 2024), complementing earlier detections made using other electromagnetic (EM) instruments (Rodriguez et al. 2006; Comerford et al. 2013; Goulding et al. 2023), and upcoming observations, by for example *Euclid* (Euclid Collaboration 2025). However, resolving the two SMBHs involved in a merger at small separations remains a significant challenge, with only a handful of resolved systems being confirmed (see e.g. Rodriguez et al. 2006; Trindade Falcão et al. 2024). This limits the insights that can be gained into SMBH mergers through EM observations alone.

The advent of GW astronomy, marked by the groundbreaking detection of merging stellar-mass BHs by the *Laser Interferometer Gravitational-Wave Observatory* (*LIGO*, Abbott et al. 2016), has revolutionized our ability to study such phenomena. Advanced *LIGO* is now pushing into the intermediate-mass BH regime, with a binary merger of mass $\sim 150 M_{\odot}$ being detected (Abbott et al. 2020). Pulsar timing arrays (PTAs) have already provided evidence for a stochastic GW background (SGWB) in the nHz range, likely sourced by a population of merging SMBHs at the high-mass end (Agazie et al. 2023b, a; Antoniadis et al. 2023; Reardon et al. 2023; Xu et al. 2023). In the next 10–15 years, the *Laser Interferometer Space Antenna* (*LISA*) will be capable of detecting SMBH mergers in the mass range of $10^4 - 10^7 M_{\odot}$ out to redshifts as high as $z \sim 20$ (Klein et al. 2016; Amaro-Seoane et al. 2017; Katz & Larson 2018). These GW observations will provide insights into SMBH merger dynamics and probe the SMBH population (Agazie et al. 2023b; Sato-Polito, Zaldarriaga & Quataert 2024; EPTA Collaboration & InPTA Collaboration 2024). Therefore, it is crucial to develop accurate theoretical models of SMBH mergers to ensure

that GW and EM data from these instruments can be interpreted effectively.

Cosmological hydrodynamical simulations have been essential in understanding the population of SMBHs in our Universe and their relation with host galaxy properties (see e.g. Sijacki et al. 2015; Habouzit et al. 2021). In anticipation of upcoming GW observations, these state-of-the-art simulations are now being used to better understand the merging population of SMBHs and make predictions for both merger rates (Kelley, Blecha & Hernquist 2016; Khan et al. 2016b; Salcido et al. 2016; Chen et al. 2022b, 2025; Li et al. 2024) and complementary EM observations of dual AGN (Rosas-Guevara et al. 2019; Volonteri et al. 2022; Chen et al. 2023a; Puerto-Sánchez et al. 2025).

Despite their varying degrees of success in explaining the properties of SMBHs, a fundamental challenge for these simulations lies in the incomplete picture of SMBH seeding mechanisms (Inayoshi, Visbal & Haiman 2020). FABLE, the simulation used in this work (Henden et al. 2018; Henden, Puchwein & Sijacki 2019, 2020), together with other widely used simulations adopt a simplistic seeding prescription, placing BH particles with an initial mass of $\sim 10^5 M_{\odot}$ at the centres of haloes that exceed a certain threshold mass (Sijacki et al. 2007; Matteo et al. 2008; Henden et al. 2018). Although recent efforts have aimed to improve these seeding models (DeGraf & Sijacki 2019; Bhowmick et al. 2024), they remain a major limitation in capturing the true nature of the formation of SMBHs and can significantly affect not just the SMBH number density but also SMBH population mass growth through the merging channel.

SMBH mergers in cosmological simulations are also hindered by numerical effects. These arise from the coarse assumptions necessary to make simulations of large cosmological volumes computationally feasible. A primary limitation is the relatively low spatial resolution in large-scale cosmological simulations, which typically causes BHs to merge at separations of several kiloparsecs, bypassing many of the important hardening mechanisms discussed earlier (Vogelsberger et al. 2013). This issue is well recognized, and to address it, many authors have applied post-processing techniques to estimate time delays at sub-kpc scales, based on the properties of the host galaxy of the remnant BH (Kelley et al. 2016, 2017; Sayeb et al. 2020; Li et al. 2022). Idealized simulations, performed at much higher spatial resolution, have also been employed to study SMBH mergers in greater detail and better constrain hardening processes at specific scales (Siwek et al. 2023; Bourne et al. 2024), although this comes at the cost of sacrificing the statistical power derived from larger volumes. More recently, significant progress has been made in incorporating these hardening mechanisms directly into cosmological simulations (Mannerkoski et al. 2021; Li et al. 2024). In addition to missing out on these important hardening mechanisms, a consequence of low spatial resolution is that BH particles might be merged prematurely when two host galaxies undergo close encounters during the galaxy merger process (Damiano et al. 2024).

Another significant limitation is the challenge of modelling DF acting on BH particles in cosmological simulations. In these simulations, the masses of surrounding DM, stellar, and gas resolution elements are often comparable to or exceed those of the BH particles, especially before the BHs have been able to significantly gain mass, leading to erratic scattering by the environment and the consequent ‘heating’ of the BH orbit (see e.g. Power et al. 2003; Ludlow et al. 2019, for a general discussion and analysis of these issues). This issue is closely related to, and further exacerbated by, the need to ‘soften’ gravitational interactions. In typical cosmological simulations, the relatively large softening length value (of the order

of $\sim 10^2 - 10^3$ pc), by definition reduces the gravitational interaction strength between the BH and surrounding matter below this scale. This effectively sets a lower limit on the minimum impact parameter that contributes to the drag force on the BH thus greatly limiting our ability to resolve DF in cosmological simulations (Merritt 1996; Athanassoula et al. 2000; Power et al. 2003; Genina, Springel & Rantala 2024). As a result, SMBH orbits in simulations are frequently unrealistic, with BHs either wandering away from the central regions of their host galaxies or failing to sink efficiently to the galactic centre. Since galactic nuclei are typically the densest regions of galaxies, these inaccuracies in BH orbits can lead to erroneous SMBH accretion rates, affecting the strength of feedback processes and, as a consequence, the environmental coupling of SMBHs (Callegari et al. 2011; Ragone-Figueroa et al. 2018). Moreover, the motion of BH particles involved in numerical SMBH mergers, even when above the spatial resolution, can remain unrealistic.

Orbit heating and wandering of BH particles are well-known numerical artefacts in cosmological simulations, and different approaches have been developed to mitigate these effects. For example, the ILLUSTRIS and ILLUSTRISTNG simulations (Genel et al. 2014; Vogelsberger et al. 2014; Sijacki et al. 2015; Weinberger et al. 2016; Pillepich et al. 2017), and similarly FABLE (Henden et al. 2018), adopt a ‘repositioning’ technique. In this method, BH particles are moved towards a local minimum of the gravitational potential within the BH smoothing length at each time-step (Springel, Di Matteo & Hernquist 2005a; Matteo et al. 2008; Sijacki et al. 2015; Ragone-Figueroa et al. 2018; Bahé et al. 2022). This approach ensures that BHs largely remain anchored to the central regions of their hosts, enabling more realistic accretion rates. However, it results in BH dynamics that are inherently unphysical. More recent efforts have focused on incorporating unresolved DF directly into simulations using sub-resolution prescriptions (Petts, Gualandris & Read 2015; Petts, Read & Gualandris 2016; Tremmel et al. 2015; Pfister et al. 2019; Bird et al. 2022; Chen et al. 2022a; Ma et al. 2023; Damiano et al. 2024). These models aim to approximate DF effects more accurately but often need to rely on idealized assumptions, such as specific velocity distributions and uniform background densities, which are rarely valid in complex astrophysical environments like galaxy mergers. Genina et al. (2024) proposed a correction for unresolved DF using calibrations based on the KETJU code, which employs post-Newtonian corrections to avoid gravitational softening in interactions of BHs and surrounding stars (Rantala et al. 2017; Mannerkoski et al. 2023). While all these methods offer unique advantages, they also have limitations. Notably, none of these approaches can accurately model DF for BH particles with masses comparable to those of DM particles, which, with present mass resolution requirements of cosmological simulations, correspond to the crucial BH seed growth and merger regime. As a result, these methods are either not yet robust enough and/or too expensive for widespread implementation in large-scale cosmological simulations, where accurately capturing BH dynamics across a range of environments is essential.

On the other hand, DF on galactic scales acting on galaxies is well modelled in cosmological simulations since it depends largely on ‘macrophysical’ properties of the galaxies, as opposed to the ‘microphysical’ properties of the BH particle and its surroundings which dictate the motion of SMBHs under DF (Boylan-Kolchin, Ma & Quataert 2007; Lotz et al. 2008). In this paper, we leverage this success to explore how numerical limitations, such as low spatial resolution and unresolved DF on BHs, influence BH merger rates in the FABLE simulation. Our focus is on the initial stage of BH hardening – namely, DF operating above the resolution limit of the simulation. We address this in post-processing by tracking the merger

process of the host galaxies within the simulation and delaying the BH merger event until the host galaxies have merged. We emphasize that these time delays arise during the well-resolved phase of galactic dynamics, and are in addition to those caused by sub-resolution hardening mechanisms. During these time delays, we also allow the BHs to accrete based on the properties of their host galaxies. We then evaluate the implications of these time delays, together with the prolonged accretion they result in for future EM and GW observations.

The rest of the paper is organized as follows. In Section 2, we describe the FABLE simulation used for this work, and validate the simulated BH population against key observational data sets. In Section 3, we turn our attention to the SMBH merger population in FABLE, investigating the dynamics of BHs during galaxy interactions, including the numerical effects leading to premature mergers, and the observational consequences. We discuss our main findings and conclude in Section 4.

2 THE BH POPULATION IN FABLE

In this work, we use the simulation suite FABLE to investigate the robustness of SMBH merger rates such simulations can predict. In this section, we will first start by analysing the population of SMBHs in FABLE. This is not only crucial to verify the validity of using these simulations to estimate the population of SMBH mergers, but is also necessary when interpreting our results in the context of present and future observations. The FABLE simulation suite is extensively described by Henden et al. (2018, 2019, 2020) and its extension to a larger cosmological volume by Bigwood et al. (2025). A brief summary is given here, with an emphasis on the modelling of BH particles in the simulation.

2.1 Methodology: the FABLE simulation

2.1.1 Basic simulation properties

The FABLE simulations employ the massively parallel code AREPO which solves the hydrodynamic equations on a Voronoi mesh in a quasi-Lagrangian fashion (Springel 2010; Pakmor et al. 2015). Gravitational interactions are modelled using the TREEPM approach whereby stars, DM and BHs are represented by collisionless particles.

FABLE employs a variety of subgrid models that simulate processes important for galaxy formation. These models are largely based on those employed in the ILLUSTRIS simulation (Vogelsberger et al. 2013, 2014; Torrey et al. 2014; Sijacki et al. 2015), including descriptions of star formation (Springel & Hernquist 2003), radiative cooling (Katz, Weinberg & Hernquist 1996; Wiersma, Schaye & Smith 2009a), and chemical enrichment (based on work by Wiersma et al. 2009b). To improve agreement with observations, the models describing stellar feedback (Vogelsberger et al. 2013) and AGN feedback (Di Matteo et al. 2005; Sijacki et al. 2007) were updated in FABLE and calibrated against the local galaxy stellar mass function and the gas mass fractions of observed massive groups and clusters (Henden et al. 2018).

The FABLE simulation suite we use in this work consists of two large-volume cosmological simulations, with periodic lengths of 40 and $100 h^{-1}$ cMpc. We will refer to these as the smaller and larger FABLE boxes, respectively, when distinguishing between the two. Since we are interested in the cosmological population of SMBH mergers, including very massive BHs relevant for the PTA regime, we primarily use the larger box in this work, hereafter simply referred to as FABLE.

This cosmological box is evolved from initial conditions based on a Planck cosmology (Planck Collaboration XIII 2016) with cosmological parameters $\Omega_\Lambda = 0.6935$, $\Omega_M = 0.3065$, $\Omega_b = 0.0483$, $\sigma_8 = 0.8154$, $n_s = 0.9681$, and $H_0 = 67.9 \text{ km s}^{-1} \text{ Mpc}^{-1} = h \times 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. The box contains 1280^3 DM particles with mass $m_{\text{DM}} = 3.4 \times 10^7 h^{-1} M_\odot$ and an initial number of 1280^3 gas resolution elements with a target mass of $\bar{m}_b = 6.4 \times 10^6 h^{-1} M_\odot$, from which stars and BHs form. The gravitational softening length is set to $2.393 h^{-1} \text{ kpc}$ in physical coordinates below $z = 5$ and in comoving coordinates for higher redshifts. The simulation is evolved until $z = 0$, with the internal parameters of all the particles and gas cells in the simulation being stored in 136 snapshots.

2.1.2 Modelling the BH population

BHs are modelled as massive, collisionless sink particles in FABLE. These particles are seeded with a mass of $10^5 h^{-1} M_\odot$ in DM haloes with a mass exceeding $5 \times 10^{10} h^{-1} M_\odot$ that do not already host a BH particle. During seeding, the highest density gas cell in the halo is converted into a BH particle (Sijacki et al. 2007; Henden et al. 2018). The smoothing length of a BH particle is defined as the comoving radius of the sphere enclosing the 64 nearest-neighbour gas cells around the BH (Nelson et al. 2019).

As in ILLUSTRIS, and many other large-scale cosmological simulations (Vogelsberger et al. 2014; Crain et al. 2015; Pillepich et al. 2017; Schaye et al. 2023), FABLE uses a repositioning scheme to ‘anchor’ the BH particles to the minimum of the host halo’s gravitational potential. At each time-step, the simulation identifies the particle with the lowest gravitational potential within a sphere of radius equal to the BH’s smoothing length, and relocates the BH particle to that position. This prevents spurious BH movement caused by two-body scattering with DM or star particles which have masses comparable to that of the BH particle, and is intended to ensure that BHs reside at the centres of host galaxies to allow BHs to accrete and grow in a realistic way (Sijacki et al. 2007, 2015; Bahé et al. 2022).

Throughout the simulation, these BH particles grow via mergers and accretion (Sijacki et al. 2007, 2015; Vogelsberger et al. 2013). Accretion happens at an Eddington-limited Bondi–Hoyle–Lyttleton-like rate

$$\dot{M}_{\text{BH}} = \min \left[\frac{4\pi\alpha G^2 M_{\text{BH}}^2 \rho}{(c_s^2 + v_{\text{BH}}^2)^{3/2}}, \dot{M}_{\text{Edd}} \right], \quad (1)$$

where ρ and c_s are the density and sound speed, respectively, of the surrounding gas, v_{BH} is the relative velocity of the BH with respect to the surrounding gas, and $\alpha = 100$ is the boosting factor (see e.g. Sijacki, Springel & Haehnelt 2009, for a discussion about α). As a consequence of the repositioning scheme described above, BH velocities in the simulation are not physically meaningful and do not represent the movement of an SMBH in a realistic astrophysical environment. Because of this, the relative velocity is neglected when calculating the accretion rate using equation (1), which clearly may lead to overestimation of the accretion rates (Vogelsberger et al. 2013). The Eddington accretion rate is given by

$$\dot{M}_{\text{Edd}} = \frac{4\pi G M_{\text{BH}} m_p}{\epsilon_r c \sigma_T}, \quad (2)$$

where m_p is the proton mass and σ_T is the Thomson cross-section. Accretion happens assuming a radiative efficiency of $\epsilon_r = 0.1$ (Yu & Tremaine 2002; Springel et al. 2005b); a fraction $(1 - \epsilon_r)$ of this accreted mass is added to the internal mass of the BH particle, while the rest is made available as feedback energy.

In certain cases, the BH particle is located in an environment with insufficient gas pressure to compress the gas and raise the density higher than the star formation threshold. In such cases, the α pre-factor that is used to account for the unresolved hot and cold phases of the subgrid interstellar medium could lead to the formation of an unphysically large and hot gas bubble around the BH accreting from a low-density medium. This leads to unrealistically high accretion rates. To correct for this, in ILLUSTRIS and FABLE, the accretion rate is multiplied by a pre-factor $(P_{\text{ext}}/P_{\text{ref}})^2$, when the pressure in the environment of the BH particle P_{ext} is less than P_{ref} , a reference pressure calculated by balancing the cooling losses in the bubble with the injected feedback energy from the BH. For further details on this pressure criterion, see Vogelsberger et al. (2013).

Feedback operates in one of two modes, and is determined by the Eddington fraction of the BH, $f_{\text{Edd}} = \dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}}$. For $f_{\text{Edd}} > \chi_{\text{radio}}$, a threshold set to 0.01 in the FABLE model, feedback occurs in quasar mode in which a fraction of the bolometric luminosity ($\epsilon_f = 0.1$) is coupled thermally to the surrounding gas and energy transfer happens isotropically (Di Matteo et al. 2005; Springel et al. 2005a). This commonly occurs at high redshift when cool gas is abundant and drives efficient BH accretion and for lower BH masses for which it is easier to reach high Eddington fractions. For $f_{\text{Edd}} < \chi_{\text{radio}}$, feedback happens in radio mode in which hot bubbles are injected into the surrounding gas at random spatial positions to mimic lobe inflation by an unresolved AGN jet (Sijacki et al. 2007). For further details on the radio and quasar modes in FABLE, see Henden et al. (2018) and Bigwood et al. (2025).

Furthermore, in Section 2.2, we study the quasar luminosity function (QLF) resulting from the BH population in FABLE. In order to calculate the bolometric luminosity of our BH particles due to accretion, we first assume that all BHs are radiatively efficient with luminosities given by

$$L_{\text{bol}} = \epsilon_r \dot{M}_{\text{BH}} c^2. \quad (3)$$

We also study the impact of a separate treatment which accounts for the radiatively inefficient AGN population, i.e. BHs accreting at low Eddington fractions. For these BHs, we use the approach outlined by Koudmani et al. (2024). This uses results from an analytic fitting function by Xie & Yuan (2012) corrected using general relativity and radiative transfer simulations (Ryan et al. 2017). Based on f_{Edd} , the radiative efficiency ϵ_r is then given by

$$\log \epsilon_r = \begin{cases} -1.0 & \text{for } f_{\text{Edd}} \geq 0.1 \\ -1.0 - \left(\frac{0.0162}{f_{\text{Edd}}}\right)^4 & \text{for } 0.023 < f_{\text{Edd}} < 0.1, \\ \sum_n a_n (\log f_{\text{Edd}})^n & \text{for } 10^{-4} < f_{\text{Edd}} \leq 0.023, \\ \sum_n b_n (\log f_{\text{Edd}})^n & \text{for } f_{\text{Edd}} \leq 10^{-4}. \end{cases} \quad (4)$$

The fitted values are $a_0 = -0.807$, $a_1 = 0.27$, $a_n = 0$ ($n \geq 2$), $b_0 = -1.749$, $b_1 = -0.267$, $b_2 = -0.07492$, and $b_n = 0$ ($n \geq 3$) (Inayoshi et al. 2019).

2.1.3 BH mergers

Two BH particles are considered to be merged in FABLE when they come within a particle smoothing length of each other. This is an adaptive length that is used to estimate hydrodynamic properties and is usually on the order of a kpc. Because of the repositioning technique used, a BH merger happens irrespective of the relative velocities of the two BHs (Sijacki et al. 2015). A merger happens instantaneously, with the remnant BH having a mass equal to the

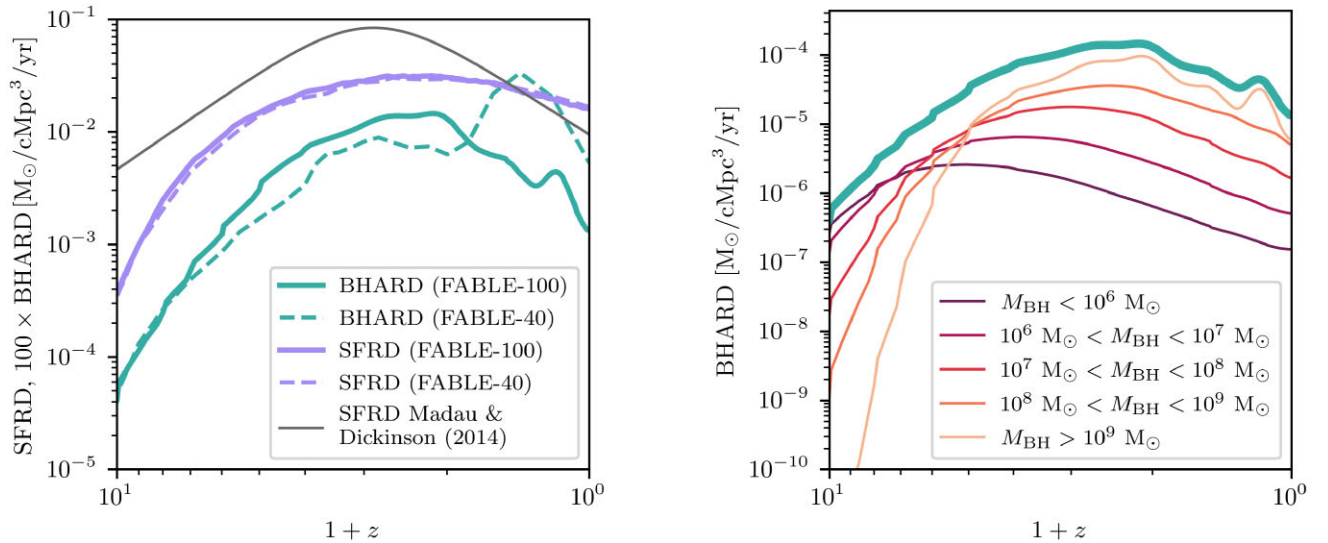


Figure 1. Left: the SFRD (purple curves) and the BHARD, scaled up by a factor of a 100 (teal curves), as a function of redshift for the FABLE simulation. The dashed lines shows results for the FABLE box with side length of $40 h^{-1}$ cMpc (FABLE-40), while the solid lines correspond to the box with side length of $100 h^{-1}$ cMpc (FABLE-100). The FABLE accretion rate densities are smoothed using a 1D Gaussian filter (see the main text for details). In grey, we show observations for the SFRD from Madau & Dickinson (2014). Right: the total BHARD (thick teal line) and the BHARD split according to BH mass (coloured lines according to the legend) as a function of redshift for the larger FABLE box. The BHARD lines are again smoothed using a 1D Gaussian filter.

sum of the masses of the merging BHs. To avoid confusion, we will refer to a merger event in the FABLE simulation as a ‘numerical BH merger’ from now on.

In addition to the information stored in the simulation snapshots mentioned in Section 2.1.1, BH specific output is also recorded at every time-step, giving better time resolution for BH accretion rates, local densities around the BHs and numerical merger events. For each merger event, the scale factor and the masses of both BHs are recorded at the exact time of the merger. After making some selection cuts, which are discussed in Section 3.2.1, this merger catalogue is used in our analysis. Note that this is the first work to use FABLE, with its more realistic galaxy groups and clusters, to study this merging population of BHs.

2.1.4 Host galaxy identification

To identify gravitationally bound structures on-the-fly in the simulation, FABLE uses the friends-of-friends (FoF) and SUBFIND algorithms (Davis et al. 1985; Springel et al. 2001; Dolag et al. 2009). The FoF algorithm first identifies ‘haloes’ by grouping together DM, stars, and gas particles using a characteristic linking length. The SUBFIND algorithm then splits these haloes into self-bound substructures called ‘subhaloes’ which can roughly be interpreted as central and satellite galaxies surrounded by their gaseous and DM haloes. This is done in configuration space by identifying density peaks. The ‘central subhalo’ is the subhalo at the minimum gravitational potential of the FoF halo, while the other subhaloes in the FoF halo are considered to be satellites.

BH particles, like other particles in the simulation, can also be associated to haloes and subhaloes. According to the BH seeding prescription used in FABLE (see Section 2.1.2), BHs are seeded in central subhaloes. However, their host subhaloes can change via a variety of physical processes, including dynamical interactions with other massive structures, or BH mergers themselves. Subhaloes might also host multiple BH particles because of these interactions. In order to study the connection between host galaxies and their

central BHs, we choose the most massive BH particle associated to a subhalo.

To track subhaloes across snaps, subhalo merger trees were constructed using the SUBLINK algorithm (Rodriguez-Gomez et al. 2015). In this algorithm, progenitors and descendants of subhaloes are identified based on their number of shared particles and their binding energies. In the context of this work, these subhalo merger trees are crucial in tracking the host subhaloes of BHs involved in a BH merger event across cosmic time, as will be explained in more detail in Section 3.1.

2.2 Results: validating the BH population in FABLE

Throughout this section, we will present the results for the BH population in the FABLE box with a side length of $100 h^{-1}$ cMpc, which is then used for BH merger predictions in Section 3, and compare this to the results from the FABLE box with side length $40 h^{-1}$ cMpc presented by Henden et al. (2018) and Koudmani, Henden & Sijacki (2021). When available, we will also draw comparisons to observational data.

The left-hand plot of Fig. 1 shows the star formation rate density (SFRD) and the BH accretion rate density (BHARD). The SFRD from FABLE is calculated by summing the total SFRD within the stellar half-mass radius of all subhaloes at a specific redshift. This is then compared to the best fit from Madau & Dickinson (2014) which uses compiled data from ultraviolet (UV) and infrared (IR) observations to obtain a comprehensive model of cosmic star formation history. Note that for a direct comparison, we multiply their fit by a factor of 0.63 to convert from a Salpeter (1955) initial mass function (IMF) to a Chabrier (2003) IMF which is what is used in FABLE. Compared to the Madau & Dickinson (2014) fit, our SFRD is lower than observations at $z \gtrsim 1$, while being slightly higher than observed values at lower redshifts. The observed SFRD peaks at $z \sim 2$, while in FABLE, the SFRD has a broad ‘high’ plateau around this redshift.

As discussed in Section 2.1.2, within the Bondi formalism, the accretion rate depends on the density and local sound speed of the gas

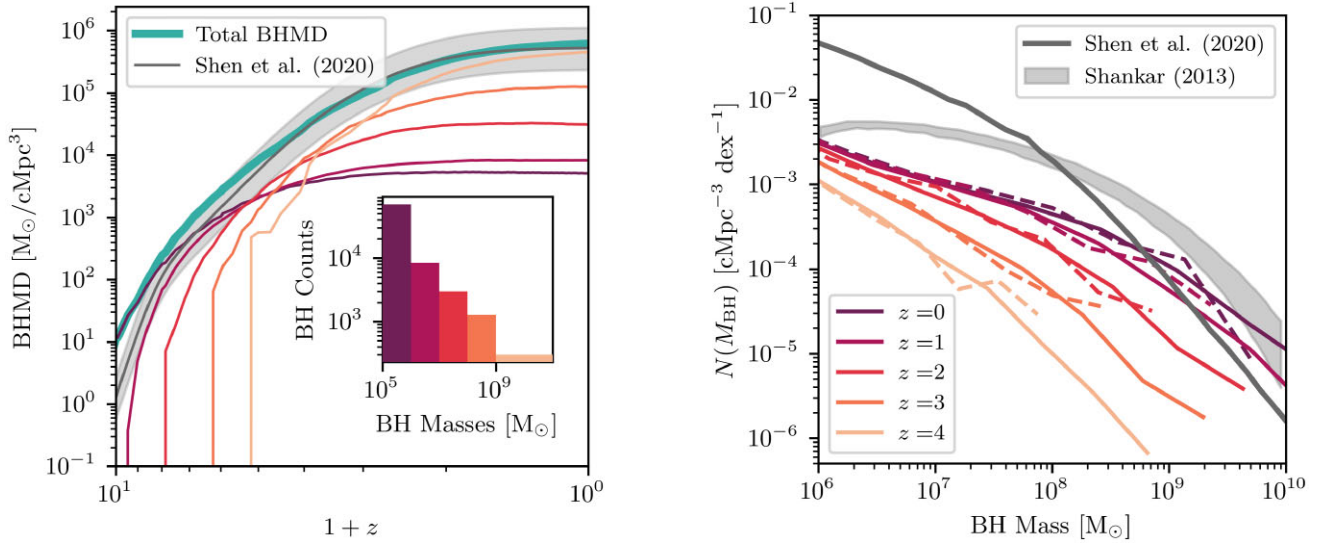


Figure 2. Left: the BHMD as a function of redshift, split according to BH mass (coloured lines are as in the legend in the right-hand plot of Fig. 1), together with the total BHMD (thick teal line) for the FABLE box with side length of $100 h^{-1} \text{ cMpc}$. The thick grey line shows an estimate by Shen et al. (2020) for the mass density of SMBHs derived from the bolometric luminosity function of AGN. This estimate assumes an average radiative efficiency $\epsilon_r = 0.1$ and a starting redshift of integration equal to 10. The grey shaded region shows the uncertainty in this estimate when increasing or decreasing ϵ_r by a factor of two. The inset plot shows the number of BHs in each mass bin at $z = 0$. Right: the BHMf at redshifts $z = 0, 1, 2, 3$, and 4, with colours as indicated by the legend, for all BHs in the FABLE simulation boxes. The solid lines correspond to the larger FABLE box, while the dashed lines correspond to the smaller box. The light grey shaded region shows the BHMf with 1σ uncertainty region calculated by Shankar (2013), assuming the $M_{\text{BH}} - \sigma$ relation defined by McConnell & Ma (2013) and applying it to all local galaxies. The thick grey line shows the BHMf calculated using the bolometric luminosity function of AGN in the local universe by Shen et al. (2020) using their ‘deconvolution’ method.

surrounding the BH (see equation 1). Prevalently, ‘radio’ mode AGN feedback causes fluctuations in the local density and temperature close to the accreting SMBHs which in turn cause the accretion rate to fluctuate. These fluctuations are smoothed using a 1D Gaussian filter in the BHARD curve so that the overall trend can be studied.¹ The differences in BHARD between the two FABLE boxes, which have the same mass and spatial resolution, is due to the differing box sizes as rarer objects corresponding to galaxy groups and clusters are less prevalent in smaller volumes. Thus, the larger FABLE box will have a larger number of these rarer objects, leading to better population statistics. In the larger FABLE box, the peak in the BHARD (around $z \sim 1.2$) is more pronounced than the peak in the SFRD and is a factor of ~ 200 lower than the SFRD. The right-hand plot of Fig. 1 shows the split of BHARD according to bins in BH mass for the larger FABLE box. The redshifts at which the BHARD for BHs at a particular mass peak reflect the cosmic time at which they contribute the most to the total BHARD with more massive BHs dominating at later times. This is due to a combination of their number density and accretion rate in absolute numbers.

The left-hand plot of Fig. 2 shows the BH mass density (BHMD) split according to different BH mass bins as indicated in the legend of Fig. 1. As expected due to their growth via accretion (right-hand plot of Fig. 1) and mergers, more massive BHs dominate the total BHMD at later redshifts. The inset plot shows the total number of BHs in each mass bin at $z = 0$. Note that as a consequence of a limited simulation volume, the most massive BHs with mass $> 10^9 M_\odot$ are scarce and hence results for this mass bin are tentative due to lack

of statistics, although the larger FABLE box offers an improvement over the smaller box as discussed above. The total BHMD is shown in the thick teal line and is compared to the estimate by Shen et al. (2020) obtained via the bolometric luminosity function of AGN observed out to $z \sim 6$ with a radiative efficiency set to $\epsilon_r = 0.1$ (solid grey line). We also show the uncertainty in this estimate when increasing or decreasing ϵ_r by a factor of two by the grey shaded region. The population of SMBHs in FABLE fits the observations very well, especially at lower redshifts.

The BH mass function (BHMf) at redshifts $z = 0, 1, 2, 3$, and 4 is shown on the right-hand side of Fig. 2. The grey shaded region shows the mass function assuming the $M_{\text{BH}} - \sigma$ relation by McConnell & Ma (2013) applied to all local galaxies with a 1σ uncertainty as calculated by Shankar (2013). The thick grey line shows the BHMf obtained from the bolometric luminosity function of AGN in the local Universe using the ‘deconvolution’ method discussed by Shen et al. (2020). FABLE underestimates the number density of BHs at all mass bins with respect to the estimates by Shankar (2013). Despite the excellent agreement with the observed BHMD, FABLE is in disagreement with the constraint on the BHMf from Shen et al. (2020), with the number of low-mass BHs being underestimated significantly (see also detailed discussion on the lack of AGN in low-mass galaxies by Koudmani et al. 2021), and the number of high-mass BHs being overestimated. This implies that although the FABLE simulation is likely correctly predicting the overall mass density of BHs, the distribution of this mass density across different mass bins is not quite in agreement with the latest observational estimates. This is driven by simplistic models of both BH seeding mechanisms and subsequent BH growth. The broader picture is similar across many widely used cosmological simulations, none of which can reproduce the full range of observational constraints (Habouzit et al. 2021).

With quasars being some of the most luminous objects in the Universe, it is crucial to check that the predicted QLF from FABLE

¹The standard deviation of the Gaussian kernel is set to the time interval between four/one simulation snapshots for the larger/smaller FABLE box. The spacing between snapshots is adaptive but is on average $\Delta a \sim 0.007$ and ~ 0.04 for the larger and smaller boxes, respectively.

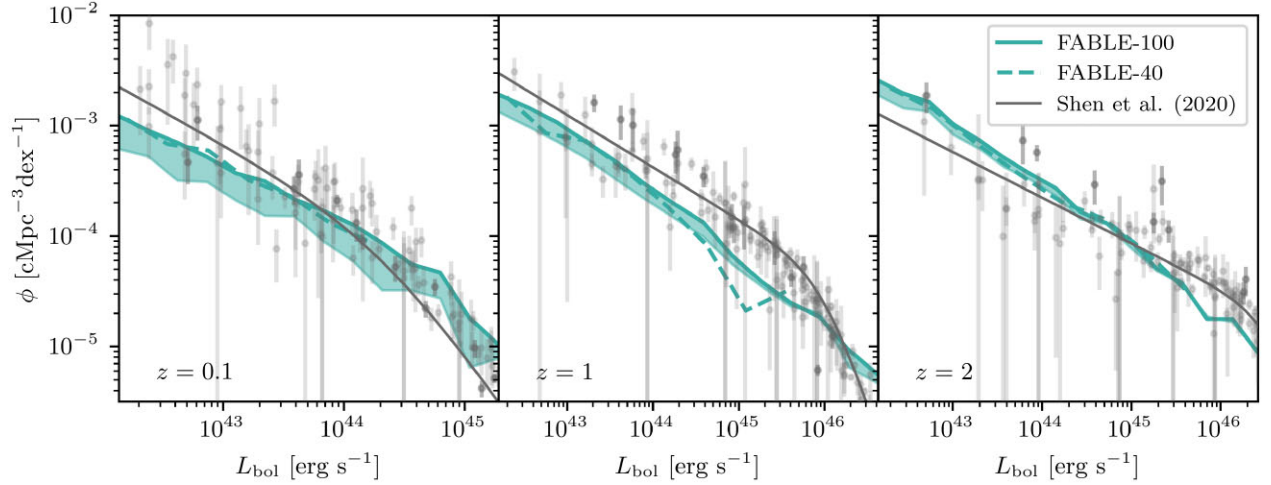


Figure 3. The AGN bolometric luminosity function at redshifts $z = 0.1$, 1 , and 2 from left to right. The solid teal lines show results from the larger FABLE box and are compared to the smaller box shown by the dashed teal lines. For the smaller FABLE box, we assume for simplicity that all BHs in the simulation are radiatively efficient. For the larger FABLE box, we show the predicted luminosity function bracketed by this assumption and that accounting for radiatively inefficient AGN as discussed in Section 2.1.2. These simulation results are compared to the global fit by Shen et al. (2020) made to observed AGN shown in the solid grey line (their global fit A is presented here). The grey points with error bars are the data used for this fit.

matches observations. However, given that the bright quasars are scarce in our simulated volumes here we focus on the entire AGN population. Fig. 3 compares the AGN luminosity function of our two FABLE boxes with the global fit to observed AGN by Shen et al. (2020) which compiles measurements made in the optical/UV, X-ray, and IR bands. We plot the results for $z = 0.1$ (for easier comparison to observational data) and $z = 1, 2$. For the simulations, we first show the results under the assumption that all AGN are radiatively efficient. Additionally, for the larger FABLE box, we provide a bracketing of the predicted luminosity function by including a population of radiatively inefficient AGN, as described in Section 2.1.2 and equation (4). Both FABLE boxes provide a good fit to the observed AGN luminosity function, especially within the context of results obtained from various cosmological simulations, as can be seen in fig. 5 of Habouzit et al. (2021). Notably, FABLE avoids the overprediction of the faint end of the luminosity function for higher redshifts which is seen in other simulations. As discussed by Bigwood et al. (2025) and Koudmani, Sijacki & Smith (2022), FABLE slightly underpredicts the QLF at the faint end at lower redshifts. This suggests that the observed SMBH population is likely accreting at more efficient rates than in FABLE in lower mass galaxies.

Fig. 4 shows the relationship between the mass of the central BH and the stellar mass within the half-mass radius, which is a simple estimate for the bulge mass, of the host subhalo. We note that using the stellar mass within the half-mass radius as a proxy for bulge mass is a common approach in cosmological simulations (see for example Matteo et al. 2008; Sijacki et al. 2015) due to the challenges of morphological decomposition when comparing to observations (Scannapieco et al. 2010; Dubois et al. 2012), though this simple approach might not capture the full complexity of bulge definitions. The best-fitting $M_{\text{BH}} - M_{\text{bulge}}$ relation by Kormendy & Ho (2013) is also shown together with data from spiral and elliptical galaxies used for this fit. While fundamental for understanding BH co-evolution with host galaxies, this mass relation has a large intrinsic scatter and becomes highly uncertain at the low-mass end due to the difficulty in observing such low-mass BHs. There is also a lack of consensus in the literature, with different authors finding somewhat different scaling relations (Kormendy & Ho 2013; McConnell & Ma 2013; Reines & Volonteri 2015; Greene, Strader & Ho 2020).

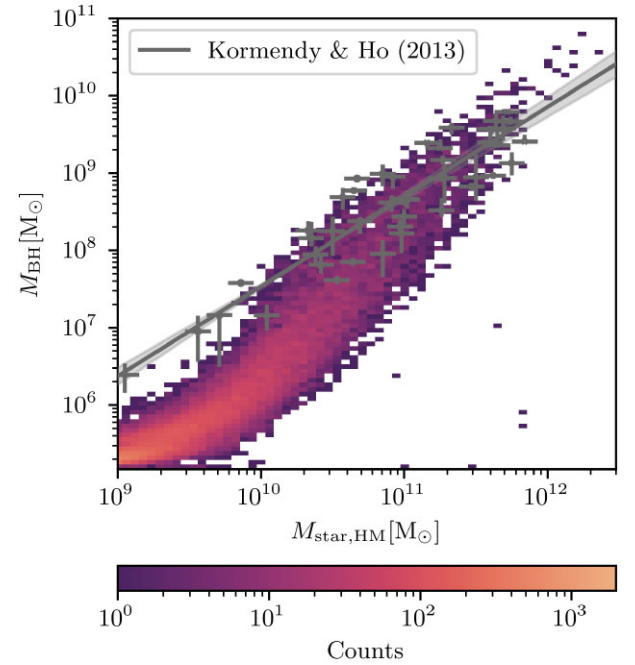


Figure 4. Central BH mass versus stellar mass within the half-mass radius ($M_{\text{star, HM}}$) of all galaxies (central subhaloes) in FABLE at $z = 0$. The thick grey line shows the best fit for the local $M_{\text{BH}} - M_{\text{bulge}}$ relation by Kormendy & Ho (2013) fitted to ellipticals and galaxies with bulges only. The grey points with error bars are the data used for this fit. Notice that SMBHs in FABLE are less massive than predicted by the local relation at the low-mass end, as also indicated by Koudmani et al. (2021).

Keeping in mind that the FABLE simulation successfully reproduces the observed galaxy stellar mass function over a wide range of masses (see fig. 1 by Bigwood et al. 2025 and fig. 2 by Henden et al. 2018), Fig. 4 shows that FABLE most likely underestimates the masses of SMBHs hosted by low (bulge) mass galaxies. This same trend is evident when considering the BH mass–stellar mass relation (evaluated within twice the stellar half-mass radius), as also concluded by

Koudmani et al. (2021), which is in line with our previous findings regarding the BHMf. Consequently, the SMBH merger rates we derive for BHs in the mass range $10^6 M_\odot \lesssim M_{\text{BH}} \lesssim 10^7 M_\odot$ should be interpreted as lower limits of the possible true rates. We investigate this issue further in Appendix A, where we plot BH mass as a function of the stellar mass within twice the stellar half-mass radius across different redshifts (Fig. A1).

3 THE MERGING POPULATION OF SMBHS IN FABLE

Building on the results in Section 2.2, we now turn our focus to the merging SMBH population in the simulation. In this section, we investigate how numerical effects lead to premature mergers within FABLE and present a first attempt at correcting for this in post-processing.

3.1 Premature mergers due to numerical effects

A BH merger event in FABLE happens as a consequence of a subhalo interaction in which both subhaloes host a BH particle. In cosmological simulations, the merging process of subhaloes is often preceded by a series of close encounters, where the two subhaloes repeatedly approach and recede from one another before the final coalescence occurs. These interactions are driven by the dynamical interplay between the gravitational potential of the two subhaloes and the surrounding large-scale environment, as well as the DF exerted by the background matter. Such interactions can also occur during subhalo fly-bys without being necessarily succeeded by a subhalo merger. During these close encounters, it is common for the two BH particles to come within a smoothing length of each other, triggering a numerical BH merger event even though the final subhalo (i.e. galaxy) merger has not yet occurred. This effect is likely often of purely numerical nature; when the gravitational potentials of the host subhaloes are sufficiently well defined for the SUBFIND algorithm to identify them as distinct structures, in many cases the BHs should remain physically bound to the centres of their respective subhaloes until the subhaloes themselves merge, if they do at all. Furthermore, recall that when two BHs are numerically merged, their relative velocities are not taken into account and there are no checks on whether they are gravitationally bound to each other. Ultimately, very high resolution simulations where the DF force is accurately followed are needed to settle this issue, but for the purpose of this work we assume that BHs should reside within the centres of their respective galaxies during the merging process as long as these galaxies can be reliably identified by the SUBFIND algorithm, which is a somewhat simplistic but much more reasonable physical scenario.

In order to study the interplay between subhalo mergers and/or interactions with BH mergers in FABLE, we identify the host subhaloes of the two BHs involved in a numerical BH merger. We do this in the most recent snapshot before the numerical merger event where both BHs reside in the central subhalo of their respective FoF haloes. When there is no common snapshot in which both BHs are in central subhaloes, we identify the host subhaloes at the earliest snapshot in which both BHs are present in the simulation. We then track these host subhaloes using the subhalo merger trees constructed using the SUBLINK algorithm as discussed in Section 2.1.4.

An example of a premature BH merger relative to the subhalo merger is illustrated in Fig. 5. This figure shows the projected gas density in a small volume surrounding the merger event, highlighting the positions of the primary BH (BH1) and the secondary (BH2), their host subhaloes, and other nearby BHs in six different simulation

snapshots. In the top-left panel, the two BHs are in the centres of their respective host subhaloes. At this cosmic time, the BHs have masses of 9.18×10^7 and $1.18 \times 10^7 M_\odot$ and reside in subhaloes with total stellar mass within twice the stellar half-mass radius of 2.66×10^{11} and $6.39 \times 10^{10} M_\odot$, respectively. The centre-top panel corresponds to the snapshot directly preceding the numerical BH merger, showing the two BHs still within their respective subhaloes, but with the secondary BH moving away from the centre of its host towards the strong gravitational potential of the primary's host. The top-right panel shows the snapshot directly succeeding the numerical BH merger, where the remnant BH is located at the centre of the subhalo that originally hosted the primary BH, while the BH2 host subhalo remains resolved as an independent structure. At this snap, the remnant BH has a mass of $1.16 \times 10^8 M_\odot$, with the stellar mass of its host galaxy being equal to $3.55 \times 10^{11} M_\odot$. The bottom-left panel shows the remnant BH undergoing a subsequent merger with another BH (the plus symbol indicates that this is no longer the direct remnant of the main merger event), and the bottom-right panel reveals the eventual coalescence of the two subhaloes into a single structure. This sequence unfolds over 11 snapshots for the selected merger event with an elapsed time of 1.3 Gyr between the numerical BH merger and the final coalescence of the two host subhaloes.

A separate but compounding numerical effect also occurs due to the repositioning scheme described in Section 2.1.2. During gravitational interactions with another subhalo, a BH particle can not only be displaced from the centre of its original host subhalo but, in some cases, also change its host subhalo entirely due to the strong gravitational influence of the interacting subhaloes. This effect is illustrated in Fig. 6 which shows a BH merger event where both BHs are removed from their original host subhaloes as a direct consequence of the repositioning scheme. At the snapshot where the host subhaloes are identified (left panel of Fig. 6), the primary and secondary BHs have masses of 1.47×10^5 and $1.60 \times 10^6 M_\odot$, respectively. They reside in subhaloes with total stellar mass in twice the stellar half-mass radius of 1.55×10^{10} and $5.44 \times 10^{10} M_\odot$, respectively. By the time of numerical BH merger, the primary BH grows to have a mass larger than the secondary. In the snapshot directly succeeding this numerical BH merger (right panel of Fig. 6), the remnant BH has mass equal to $7.71 \times 10^8 M_\odot$ and resides in the central subhalo of the FoF halo with total mass equal to $1.01 \times 10^{12} M_\odot$. The original host subhaloes have masses of the order of $10^{10} M_\odot$ and are substructures within the larger central subhalo whose strong gravitational potential completely displaces the remnant from the original satellite subhaloes hosting the BHs. Similarly to the premature mergers during subhalo interactions depicted in Fig. 5, this displacement of BHs away from their original host subhaloes is purely a numerical artefact. Physically, BHs are expected to remain anchored at the centres of their host galaxies during interactions with passing galaxies, as long as the innermost dense region to which the BH is bound to exists, except in cases where those galaxies themselves merge. In some instances, the original host subhaloes of the merging BHs can go on to evolve independently, with a large spatial separation persisting between them. This can result in the original subhaloes even never merging at all, despite their BHs having merged within FABLE.

3.2 Methodology: macrophysical time delays to SMBH mergers

In order to quantify the numerical effects described above, we introduce a time delay in post-processing corresponding to the interval between the numerical BH merger and the potential eventual merger of their host subhaloes. It is important to stress that these

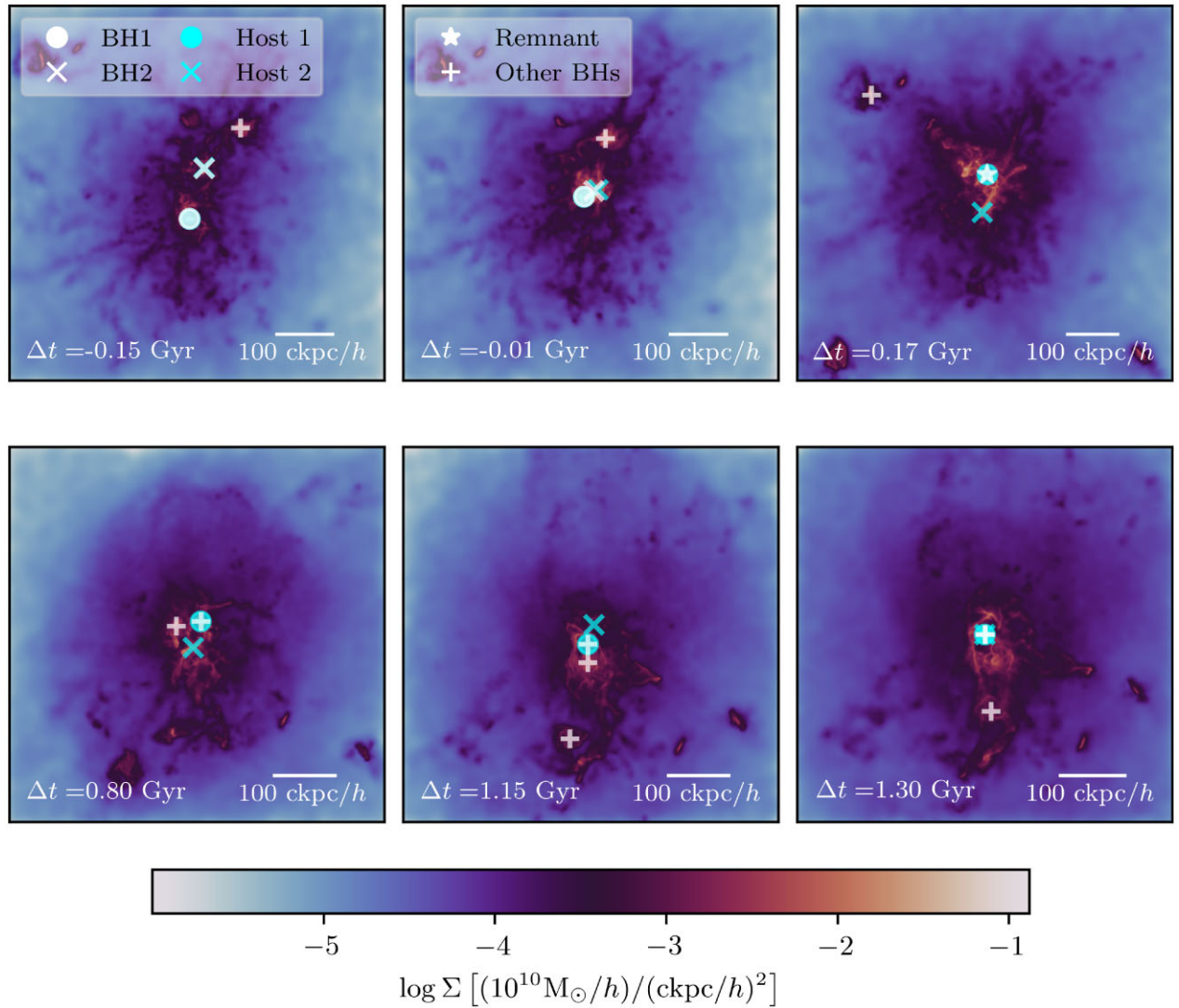


Figure 5. The projected gas density in a small cosmological volume surrounding a representative merger event in FABLE for six different snapshots in the simulation. The positions of the primary (BH1) and secondary (BH2) BHs are shown as a white circle and cross, respectively. The positions of their respective host subhaloes are shown using the same marker shapes, but in cyan. The direct remnant of the BH merger of interest is shown as a white star, when present, while the positions of any other BHs in the vicinity are shown with a white plus symbol. The top-left panel shows the BHs two snapshots before the numerical BH merger, with both BHs residing at the centre of their respective host subhaloes. The top-centre panel corresponds to the snapshot directly preceding the numerical BH merger and shows BH2 moving away from the centre of its host subhalo due to the repositioning technique used in FABLE. The top-right panel shows the snapshot succeeding the merger, with the remnant BH residing in host subhalo that hosted BH1 initially, and with the two subhaloes still resolved as two separate structures. The bottom-left panel shows the remnant BH merging with another BH particle (notice the change from a star marker to a plus marker), while on the bottom-right panel, we see the two host subhaloes finally merging into one subhalo. Δt in the bottom-left corner of each panel corresponds to the time elapsed since the numerical BH merger, with negative values indicating time to numerical BH merger. The scale in the bottom-right corner of each plot is given in simulation units.

macrophysical time delays are distinct from, and in addition to, the microphysical time delays extensively studied in the literature (see the Introduction). The latter are usually calculated from the time of numerical BH merger and account for processes governed by unresolved subresolution physics.

3.2.1 Selection cuts

The macrophysical time delays described above are applied to the BH merger population in FABLE after a few selection cuts are made. As seen in Fig. 2, the distribution of BH masses peaks at the low-mass end. These low-mass BHs are usually located in satellite subhaloes

which quickly merge with a nearby, more massive subhalo, leading to a numerical BH merger. Because of our poor understanding of seeding mechanisms we only keep merger events in which both BH particles have a mass $M_{\text{BH}} \geq 10^6 M_{\odot}$ at the time of numerical BH merger. There are a total of 91 879 numerical BH mergers in FABLE out of which 11 016 (~ 12 per cent) involve two non-low-mass BHs.

Our method assumes that DF on galactic scales is well resolved, ensuring that the subhalo merger time-scales in cosmological simulations are reliable (Rodríguez-Gomez et al. 2015). To mitigate numerical artefacts that could undermine this assumption, we focus exclusively on merger events where the constituent BHs reside in subhaloes with a total stellar mass, measured within twice their

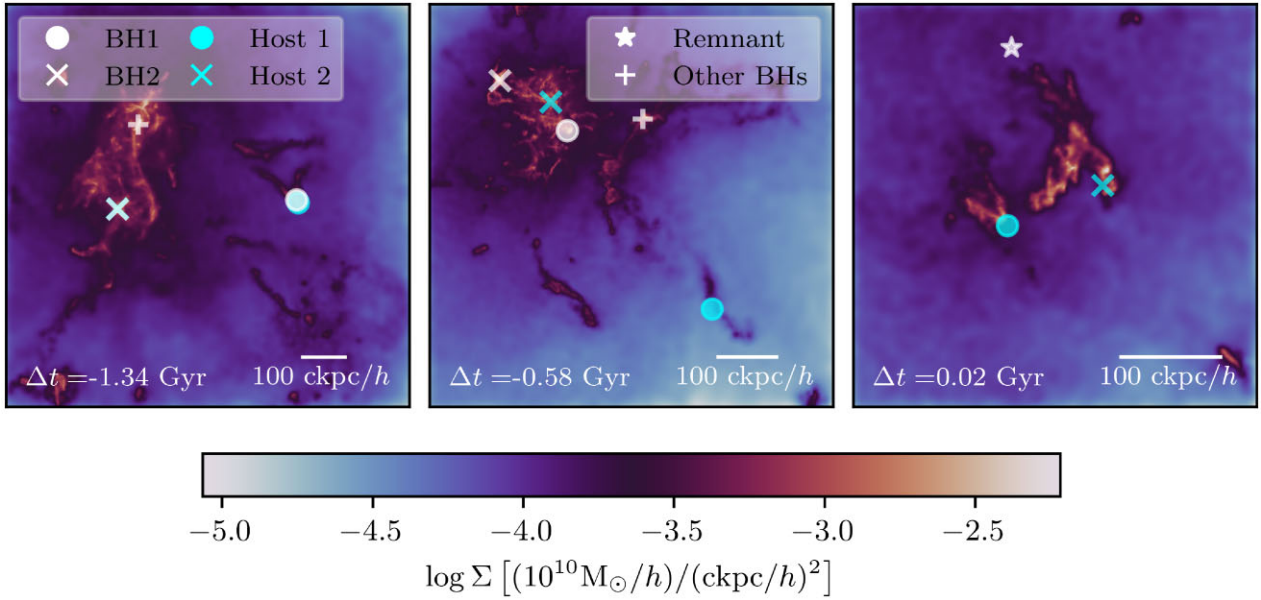


Figure 6. The projected gas density around a second representative BH merger event in FABLE. The legend is the same as described in the caption of Fig. 5. The left panel shows the BHs residing at the centres of their respective host haloes. The centre panel shows the two BHs residing in completely different subhaloes than their original hosts, while the right panel corresponds to the snapshot directly succeeding the numerical merger event, with the remnant BH residing in neither of the original host subhaloes. The original host subhaloes do not merge by $z = 0$ in FABLE.

stellar half-mass radius, of at least 100 times the mass of a star particle, i.e. $100 \times (6.4 \times 10^6 h^{-1} M_\odot)$, at the time of host subhalo identification. This further reduces our sample size from 11 016 down to 10 716 BH merger events.

3.2.2 BH accretion

If BH particles were to remain at the centres of their host subhaloes between their numerical merger and the eventual subhalo merger, they would continue accreting mass, leading to more massive BHs at the time of merger. To account for this additional growth, we incorporate simple accretion onto the BHs in post-processing.

To do this, we replicate the simulation’s accretion model as described in Section 2.1.2. At each snapshot between the numerical BH and the subhalo mergers, we retrieve the physical properties of the 32 gas cells closest to the position of the subhalo’s most bound particle. These properties are used to calculate the density, sound speed, and pressure in the central region where the BH particle would reside, with values interpolated for the time intervals between snapshots. The BH mass is then evolved according to the Eddington-limited Bondi–Hoyle–Lyttelton rate, as defined in equation (1).² Note that while this approach allows us to gain insight into the likely BH growth between numerical BH and subhalo mergers, by construction this post-processing model cannot account for BH feedback which may likely affect the central subhalo properties and the rate of BH accretion itself, an issue both for the original FABLE simulation and our post-processing analysis.

²Although the v_{BH} term in the Bondi–Hoyle–Lyttelton formula can suppress the BH accretion rate, we omit it, as is done in FABLE and many other cosmological simulations that implement BH repositioning (see Section 2.1.2), in order to replicate the simulation’s accretion prescription as closely as possible. We also apply the same pressure criterion as described in Section 2.1.2.

3.2.3 Intervening mergers

Mergers between two BH particles in FABLE do not happen in isolation. It is common for such merger events to occur in close proximity to other mergers within a relatively short period of time, for example, when multiple satellite subhaloes, each hosting central SMBHs, merge with a central subhalo hosting another SMBH. Removing low-mass BHs from our sample, as discussed in Section 3.2.1, reduces the number of such events drastically, although it does not completely eliminate them.

Incorporating macrophysical time delays can potentially alter the sequence of merger events in the simulation. This effect is illustrated in Fig. 5, where a third BH merges with the remnant BH from the initial merger before the host subhaloes have fully merged. If we were to recompute the order of the mergers within our model, this third BH should merge with BH1 before the remnant of this event then merges with BH2. Such re-ordering could influence the total masses and mass ratios of the BHs involved in these mergers. However, only 452 out of the 10 716 merger events (4 per cent) that meet our selection criteria experience such a sequence reversal due to macrophysical delays. While other events may involve intervening mergers with low-mass BHs, these are not expected to notably affect the mass ratios or chirp masses of the primary merger events. Given the low frequency of intervening mergers with non-low-mass BHs, we neglect this effect without significantly compromising the population-level results.

3.3 Results: macrophysical time delays to SMBH mergers

In this section, we analyse the macrophysical time delays discussed in Section 3.2 to quantify the impact of numerical effects that lead to premature BH mergers within FABLE.

From a total of 10 716 mergers which pass our selection criteria as described in Section 3.2.1, only 513 merger events have no added time delay (~ 5 per cent). In these merger events, the host subhaloes merge before the BH particles, i.e. the BH merger happens within

the remnant host galaxy. These 513 events occur across a range of different redshifts and involve different BH masses. Although in these events, there is a trend towards more massive primary BHs of order $10^{10} M_{\odot}$ residing in haloes at the more massive end of the mass spectrum, and secondary BHs residing in more massive subhaloes. This is consistent with expectations that more massive subhaloes will have a deeper gravitational potential and thus are more capable of ‘holding on’ to their central BHs during subhalo mergers. In such cases, after the subhalo merger, the BH orbits are artificially influenced by the repositioning scheme, and microphysical delays should therefore be applied starting from the earlier time of subhalo merger rather than numerical BH merger. However, this correction would only affect a small fraction of the BH merger population and we do not model it in this work. In 541 merger events, the host galaxies merge in the snapshot directly after the numerical BH merger.

Our time-delay estimates are limited by the coarse temporal resolution of subhalo data, which is recorded only at snapshot intervals, unlike the more finely sampled BH merger data. As a result, delays are slightly overestimated, since the actual subhalo merger could occur at any time between the last snapshot where both subhaloes are resolved and the subsequent one. The snapshot spacing varies with redshift, averaging ~ 0.1 Gyr at $z = 3$ and ~ 0.15 Gyr at $z = 0$, which sets an upper bound on the timing uncertainty.

In some cases, the host galaxies of the BHs involved in a merger event fail to merge by the end of the simulation, i.e. by $z = 0$. This happens for 3140 events (~ 29 per cent). For brevity, we will refer to such events as ‘non-mergers’ from here onwards.

The distributions of macrophysical time delays in our model, with different treatments of non-mergers, are shown in Fig. 7. The pink histogram shows the distribution of time delays assuming that all host galaxies merge by redshift zero, i.e. for non-mergers, we calculate the time elapsed until redshift zero from the point of numerical BH merger. We then omit these non-merger events and show the macrophysical time-delay distribution for those events whose host galaxies do merge by the end of the simulation, in teal. We use this distribution to obtain a probability distribution function and use this to estimate the total macrophysical time delay for non-mergers based on the time elapsed between the numerical BH merger and $z = 0$. This distribution is then shown by the purple line. This latter distribution gives a maximum macrophysical time delay of 12.6 Gyr and a median of 1.3 Gyr. We find a weak correlation between specific gas mass fraction and our macrophysical time delays, with shorter delays associated with higher gas fractions. This likely arises because, in denser environments, the BHs have smaller smoothing lengths, reducing the likelihood of numerical BH mergers during subhalo interactions.

The merger event illustrated in Fig. 5 has a time delay equal to 1.3 Gyr, corresponding to the median value in our sample. This delay, indicated by the vertical purple dashed line in Fig. 7, provides a representative example of the merger events in our sample. Notably, the host galaxies in Fig. 5 undergo significant morphological transformations between the numerical BH merger and the coalescence of the host subhaloes. The distribution of time delays in Fig. 7 thus suggests that these macrophysical delays drive substantial changes in the morphology of SMBH merger host galaxies. To compare to microphysical delays obtained from cosmological simulations, we refer to the study by Kelley et al. (2016) in which the hardening time-scales from DF, stellar scattering, gas drag, and GW emission are modelled using a post-processing approach based on the properties of the remnant host galaxy. In their fiducial model, they find a median microphysical time delay for the full sample of 29 Gyr.

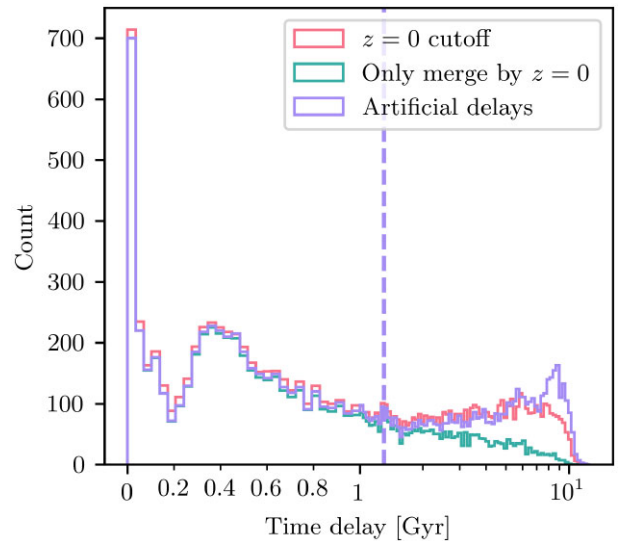


Figure 7. Distribution of macrophysical time delays in our post-processing model. The pink distribution assumes that all host galaxies merge by the end of the simulation. This is done by calculating the elapsed time until $z = 0$ for merger events whose host galaxies do not merge within the simulation. The teal distribution removes these events and only shows the time delays of BHs residing in host galaxies that merge by $z = 0$. The purple curve shows the time delays after artificially delaying those BH mergers whose host galaxies do not merge by $z = 0$ as described in Section 3.3. Time delays until 1 Gyr are shown on a linear scale, while longer time delays are shown on a logarithmic scale. The vertical dashed line shows the time delay of 1.3 Gyr for the merger event shown in Fig. 5 and corresponding to the median of the latter distribution.

This is much longer than the age of the Universe and only 20 per cent of the binaries coalesce before $z = 0$. The median is reduced to 6.9 Gyr when considering only systems with a total mass $> 10^8 M_{\odot}$ and mass ratio > 0.2 . Although these values strongly depend on model assumptions such as the loss-cone refilling rate, they are comparable to our macrophysical delays, highlighting the importance of considering both effects in predictions for future observations.

The suppression between the first two distributions (pink and teal) in Fig. 7 is directly caused by non-merger events, by definition of the distributions. While a slight suppression is present across the entire range of time delays, it becomes most pronounced for delays exceeding ~ 1 Gyr. The consistent suppression across the full time range corresponds to premature BH mergers during ongoing galaxy mergers, as highlighted in Fig. 5, that do not merge by $z = 0$, resulting in non-merger events. This effect is particularly relevant at lower redshifts. However, the majority of these non-mergers, which dominate the tail ends of the distributions, stem from the numerical effect illustrated in Fig. 6 in which BH particles are repositioned to different subhaloes and lose their association with their original host subhaloes. Consequently, these hosts often evolve independently, and are likely to never merge, leading to large macrophysical delays in our model.

To calculate time delays from subgrid hardening mechanisms, as described in the Introduction, an initial separation at which to initialize these subresolution prescriptions is required. This should roughly match the separation of the BHs at which their orbits within the cosmological simulation are no longer tracked, i.e. the separation at which they are considered to be merged within the simulation. In previous works, an upper limit for this initial separation was taken to be the smoothing length of the more massive BH particle in the snapshot directly preceding the numerical merger event (Kelley

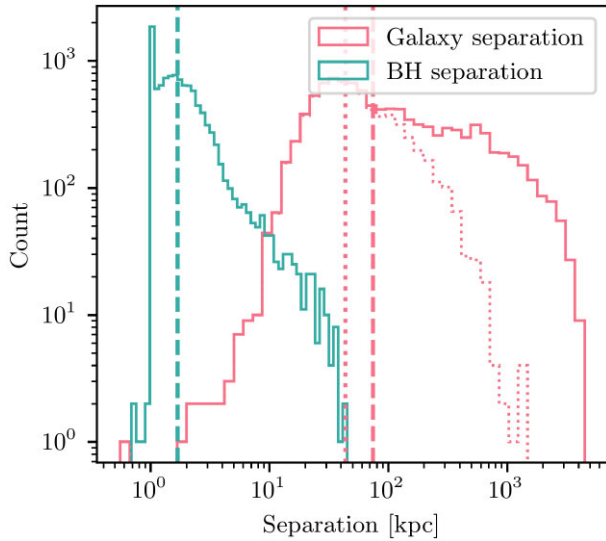


Figure 8. Separations of BHs calculated using the smoothing length of the more massive BH in teal, compared to the host galaxy separations in the snapshot preceding the host–galaxy merger in pink. The solid pink distribution includes the separation of the host subhaloes at $z = 0$ for non-mergers. The dotted pink line plots the distribution of host galaxy separations without these non-merger events. The vertical lines show the median of the distributions, with the dashed lines corresponding to the two former distributions (in matching colours) and the dotted line corresponding to the events without non-mergers.

et al. 2016), or a small multiple of the spatial resolution of the simulation (Li et al. 2024). We plot the distribution of BH separations using the particle smoothing lengths for our sample of merger events in FABLE in teal in Fig. 8, which matches distributions shown by Kelley et al. (2016). However, since the time of BH merger in FABLE is dominated by numerical effects, as depicted in Figs 5 and 6, the corresponding BH separations at merger are also plagued by these numerical artefacts, and no longer correspond to the particle smoothing lengths at the redshift at which we consider the BHs to be finally merged within our macrophysical delay model.

We argue that a better, albeit more conservative estimate for the initial separation which can be obtained from these large-volume cosmological boxes, is the separation between the original host subhaloes, in the snapshot before subhalo merger. The distribution of these separations for the same sample of merger events is shown in pink in Fig. 8, with the solid line corresponding to the full merger sample and the dotted line excluding non-merger events. The corresponding vertical dashed lines in Fig. 8 indicate the median for each distribution, with the median galaxy separation being roughly two orders of magnitude larger than the BH separation set by their smoothing lengths. This would imply that a longer phase of orbit shrinking via DF is required when modelling subresolution hardening mechanisms using the galaxy separation as an educated guess for the initial separation. This would inadvertently lead to significantly longer microphysical time delays than our current best estimates when using BH smoothing lengths, compounding the added macrophysical delays from our model.

3.3.1 Implications for GW observations

In order to quantify the effects of our macrophysical delays on GW observations, we compare the subsample of merger events passing our selection criteria in FABLE without any modifications, with added

macrophysical time delays but no modification to their masses, and with added delays while allowing for growth by accretion to happen as discussed in Section 3.2.2. In order to isolate the effects of our macrophysical delays on the observable population of SMBH mergers, we neglect any microphysical time delays arising from unresolved hardening mechanisms, unless otherwise stated.

As indicated in Fig. 7, in our model the number of merger events is reduced, as some host galaxies do not merge by the end of the simulation. Consequently, introducing macrophysical delays alters the mass distribution of the merger population, with accretion further modifying their masses. Fig. 9 displays the number density of merger events occurring before $z = 0$ as a function of key mass parameters: primary mass M_1 , secondary mass M_2 , chirp mass \mathcal{M} , and mass ratio $q = M_2/M_1$, where the chirp mass is defined as

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

The distribution with respect to primary mass (top-left) shows that while the number of mergers decreases across all masses after introducing macrophysical delays, the suppression is strongest at high M_1 , which is then reflected in the chirp mass distribution (bottom-left). This effect arises because SMBH growth through accretion and mergers leads to the formation of larger BHs at later cosmic times in FABLE, as highlighted by the BHMF in Fig. 2. Since high-mass mergers predominantly occur at lower redshifts, applying time delays shifts many of these mergers into the future, removing them from the observed population. In contrast, the suppression in secondary mass M_2 (top-right) is more uniform across the full mass range. The reduction in high-mass primaries also decreases the number of mergers with the most extreme mass ratios, though the overall number of merger events declines across all q values (bottom-right). Notably, incorporating accretion during the added macrophysical delays does not significantly alter any of the distributions in Fig. 9.

We further investigate this accreted mass during our macrophysical delays in Fig. 10. This figure shows the distribution of normalized accreted mass for merger events whose host galaxies merge by $z = 0$. Normalized accreted mass is defined as $(M_{\text{BH,merger}+\Delta t_{\text{macro}}} - M_{\text{BH,merger}})/M_{\text{BH,merger}}$, where $M_{\text{BH,merger}}$ is the BH mass at the time of numerical BH merger, while $M_{\text{BH,merger}+\Delta t_{\text{macro}}}$ is the BH mass after the macrophysical time delay. We exclude BHs which accrete no mass during the macrophysical time delay, either due to this delay being equal to zero, or because they reside in gas-poor galaxies. We find that the amount of mass accreted during this delay is generally low. However, about 4 per cent of the BHs in this figure at least double their mass, likely corresponding to the tail of the macrophysical time-delay distribution (teal colour), showing BHs that merge by $z = 0$, in Fig. 7. We also find that secondary BHs systematically accrete more mass than primary BHs during these delays. As discussed in Section 3.2.2, this happens because our post-processing approach cannot self-consistently model BH feedback. Secondary host galaxies which are prematurely stripped of their central BH particle (see Fig. 5) will not have any BH feedback to regulate central gas densities, allowing BHs to accrete at higher rates in post-processing. On the other hand, primary host galaxies which tend to host the more massive remnant BH during this delay than should be the case, experience feedback that is too strong (due to BH mass dependence), pushing gas away from the central regions of the galaxy, leading to lower accretion rates calculated in post-processing. This might also affect the SFRs in these host galaxies.

Fig. 11 illustrates the relation between merger chirp mass \mathcal{M} and redshift z in FABLE for our three different models. Note that the teal contours, corresponding to the unmodified population in FABLE,

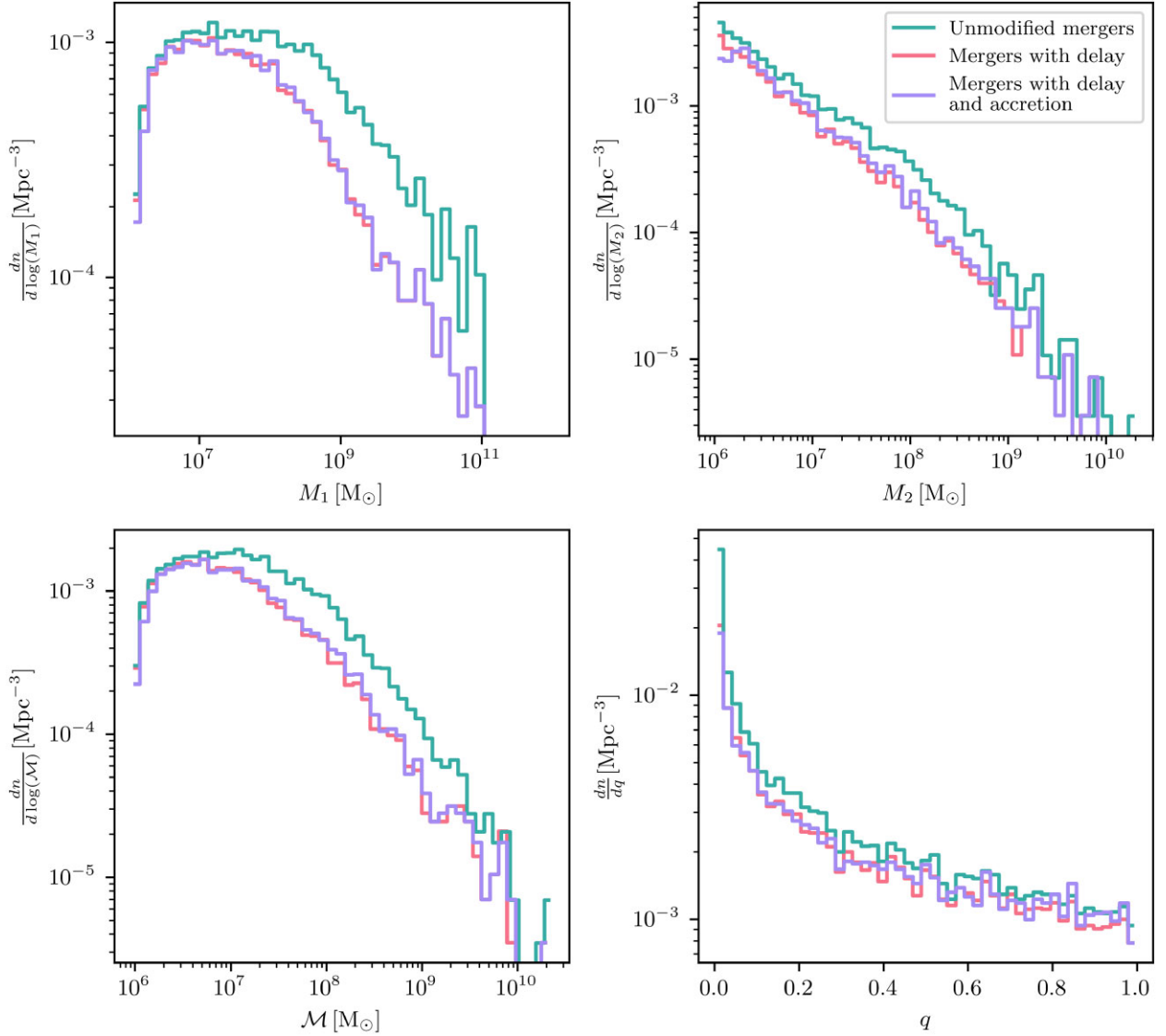


Figure 9. Distributions of the number density of events as a function of primary mass M_1 (top left), secondary mass M_2 (top right), chirp mass \mathcal{M} (lower left), and mass ratio q (lower right). The teal lines correspond to the unmodified merger population in FABLE, the pink lines add our macrophysical delays without changing the BH masses, and the purple lines show the distribution when accounting for BH accretion during the macrophysical delays.

contain a larger number of events than the contours including delays, as the latter do not include non-merger events, making them a subset of the original population. In the top and right subpanels, the marginal distributions of chirp mass and redshift, respectively, are shown.

Our macrophysical delays slightly suppress events at redshifts $z > 1$, with the effect of increasing the number of events occurring at $0 < z < 1$, although the location of the peak of the redshift distribution does not shift significantly. A more pronounced effect is the shift toward lower chirp masses across all redshifts which can be seen both in the 2D and 1D distributions. The broad peak in the distribution of chirp masses at $\sim 10^7 M_\odot$ produced by FABLE is shifted towards $\sim 3 \times 10^6 M_\odot$ and becomes more pronounced. The teal contours from the unmodified FABLE simulation in the 2D contour plot highlight how high-mass mergers occur at later redshifts. Showing the joint distribution of chirp mass and redshift breaks the degeneracy seen in the top panel and in Fig. 9 which group mergers across all redshifts. Because of this, it is clearer to see the impact of accretion on the distribution of chirp masses at specific redshifts. In general, our

accretion model shifts mergers at fixed z -values towards higher chirp masses, however this effect is not large enough to recover the highest chirp mass events that are lost with the introduction of macrophysical time delays.

In the shaded, grey background of Fig. 11, we plot the contours of constant *LISA* signal-to-noise ratio (SNR), i.e. the ‘waterfall plot’. The SNR calculations were performed using the tool Binary Observability With Illustrative Exploration (BOWIE, Katz & Larson 2018) and the ‘IMRPhenomD’ waveform (Husa et al. 2016; Khan et al. 2016a). For these SNR calculations, we assume two non-spinning BHs with a fixed mass ratio of $q = 0.5$ on a circular orbit and use the sky-averaged *LISA* phase A noise curve above 10^{-4} Hz, including the confusion noise from galactic white dwarf binaries (Ruiter et al. 2010; Colpi et al. 2024). Despite these simplifications, these sensitivity curves give a good impression of *LISA*’s sensitivity and show the implications of our macrophysical time delays. For the high-mass binaries studied here, *LISA* will observe the entire GW signal above 10^{-4} Hz which lasts $\lesssim 1$ month.

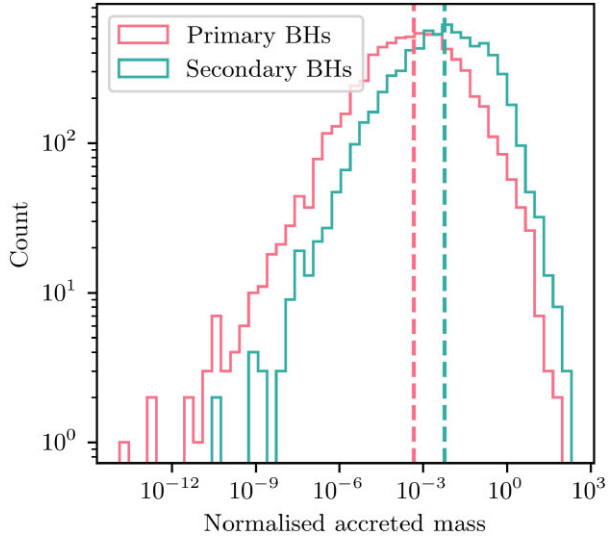


Figure 10. Normalized accreted mass between the numerical BH merger and the end of the macrophysical time delay, i.e. $(M_{\text{BH,merger}} + \Delta t_{\text{macro}} - M_{\text{BH,merger}})/M_{\text{BH,merger}}$ (see the main text for more details), for the subset of BH merger events whose host galaxies merge by $z = 0$. The distribution for primary BHs is shown in pink, while that for the secondary BHs is shown in teal. The corresponding vertical dashed lines indicate the median of each distribution. This shows that secondary BHs systematically accrete more mass than primary BHs in our post-processing approach.

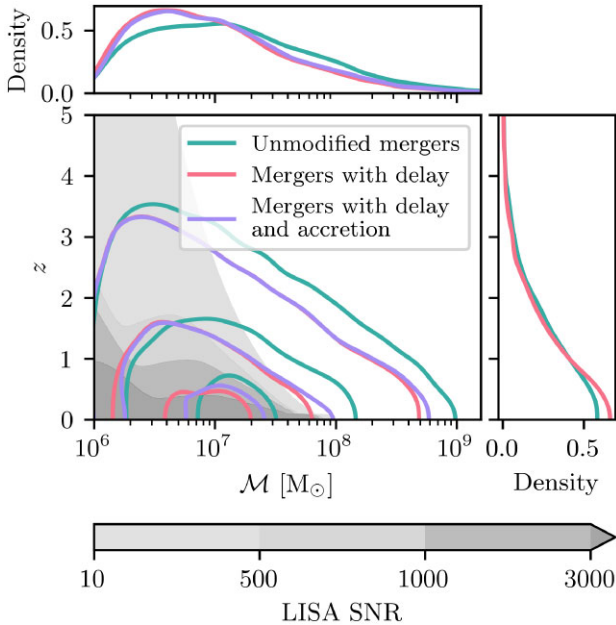


Figure 11. The main figure shows the distribution of merger events in redshift-chirp mass space. The teal contours show this distribution from FABLE, the distribution shown in pink adds our macrophysical time delays in post-processing whilst leaving the BH masses unchanged, and the distribution shown in purple adds the same macrophysical time delays whilst also allowing the BHs to accrete during this delay time. The contours show 10 per cent, 50 per cent, and 90 per cent of our sample. Grey shaded regions show the curves of constant SNR for *LISA* as indicated by the colour bar and as discussed in Section 3.3.1. The top plot shows the marginal distribution of chirp masses, while the right-hand side plot shows the distribution of redshifts.

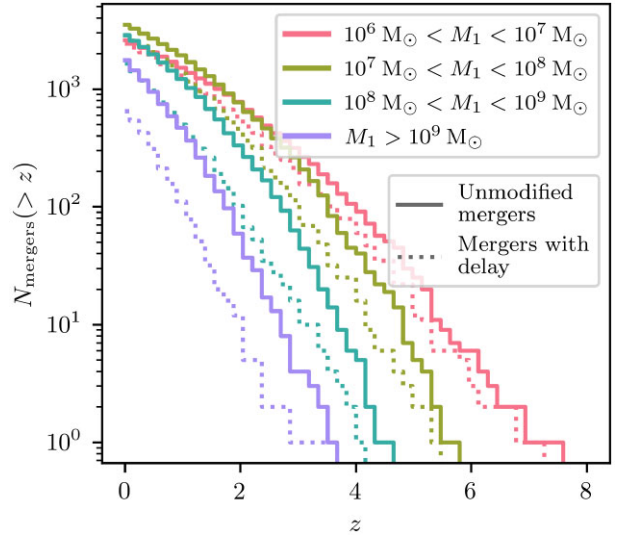


Figure 12. Cumulative merger number binned by primary mass. The solid lines show the unmodified merger population in FABLE, whilst the dotted lines show the cumulative distribution after adding macrophysical time delays.

LISA will also observe SMBH mergers with smaller chirp masses which are not included here because they do not pass the mass cut applied to the FABLE catalogue described in Section 3.2.1. The shift in the SMBH population towards lower redshifts due to macrophysical delays means that the binaries that merge will have higher *LISA* SNRs than predicted by FABLE, improving their detectability. Furthermore, the suppression of high-mass mergers implies that *LISA* will be able to detect a larger percentage of merger events at higher SNR, leading to a more complete understanding of the merging SMBH population.

The suppression of high-mass merger events is also highlighted in Fig. 12 which shows the cumulative number of mergers as a function of redshift binned according to the primary mass, M_1 . The unmodified merger population in FABLE is compared to the merger population after the addition of our macrophysical time delays. Note that accounting for accretion does not significantly alter the BH masses, and therefore does not affect the binning or the results shown in this figure. The number of mergers is reduced across all mass bins with the introduction of these time delays although the effect becomes more noticeable as the mass of the primary increases, in agreement with Figs 9 and 11. In both cases, BH mergers with $10^7 M_\odot < M_1 < 10^8 M_\odot$ dominate the merger rate which is broadly consistent with previous predictions from cosmological simulations (Salcido et al. 2016; DeGraf & Sijacki 2019). At higher redshifts, the number of mergers increases with decreasing primary mass which is consistent with the cosmic growth of SMBHs via accretion and mergers.

Fig. 13 shows the expected SGWB sourced by the population of merger events in FABLE passing our selection criteria within the PTA frequency sensitivity range for all three of our models. To calculate the strain, we use the HOLODECK package for SMBH binary population synthesis and GW calculations (Kelley et al. 2016; Agazie et al. 2023b). Based on each population of mergers, discrete realizations of the binary population are generated using a weighted sampling approach, with each SMBH binary in our sample being weighted using a Poisson distribution. For each realization, the total strain in each observed frequency bin is calculated by averaging over the angle and polarization for binaries on circular orbits (Finn & Thorne 2000). Accounting for unresolved hardening processes is expected to significantly further delay SMBH mergers, but their

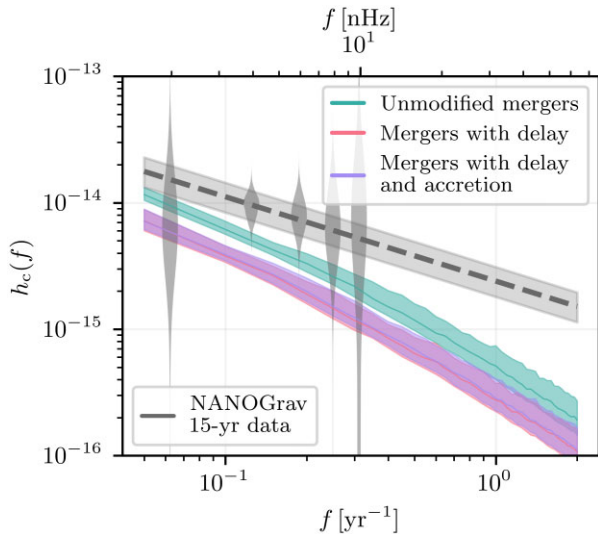


Figure 13. The SGWB strain generated from the subsample of merger events in FABLE passing our selection criteria (see Section 3.2.1). The unmodified population is shown in teal, the population with added macrophysical time delays is shown in pink, while the population with added macrophysical time delays as well as allowing for BH accretion during these delays is shown in purple. For all simulation models we further assume that unresolved hardening microphysical delays are equal to 1 Gyr. For the unmodified merger population we use the initial separations set by the BH smoothing lengths whereas for the two populations with added macrophysical delays we initialise the binaries at the separation of the host galaxies. We draw 1000 realizations from our discrete binary population in order to estimate the SGWB in each case. The solid lines indicate the median of these realizations, while the interquartile range is highlighted by the shaded regions. The grey dashed line with the shaded region shows the power-law fit from the NANOGrav 15-yr data set (median with 90 percent credible interval). The grey violins show the Hellings–Downs correlated free spectrum posteriors from the same data set for the first five frequency bins (see Agazie et al. 2023b, for more details). The frequency range plotted corresponds to the sensitivity of PTAs after 20 yr of observations.

modelling still involves large uncertainties. To isolate the effect of our macrophysical delays, we adopt a fixed microphysical hardening time delay of 1 Gyr as a simplifying assumption, following the approach by Agazie et al. (2023b). We set the initial separations of the unmodified merger events to the particle smoothing lengths of the primary BHs (teal distribution in Fig. 8). For the two populations with added macrophysical delays we use the host galaxy separations (pink distribution in Fig. 8) as the initial separations.

As expected, the addition of macrophysical delays suppresses the SGWB across all frequencies due to the reduction in the number of merger events that occur before $z = 0$, specifically at the high-mass end of the distribution, as illustrated in Fig. 12, which is the main contribution to the SGWB. Allowing the BHs to grow via accretion during the added macrophysical delays does not significantly boost the amplitude of the signal. Theoretically, for a large population of sources, and assuming circular orbits decaying primarily due to GW emission, the characteristic strain is expected to scale with frequency as $h_c(f) = A_{\text{GWB}} (f/f_0)^{-2/3}$ (Phinney 2001). We show this as the grey dashed line in Fig. 13 using the fiducial result from the North American Nanohertz Observatory for Gravitational Waves (NANOGrav) power-law analysis of the 15-yr data set with an amplitude at the reference frequency of $A_{\text{GWB}} = 2.4^{+0.7}_{-0.6} \times 10^{15}$. We also show the results of the Hellings–Downs correlated free-spectrum analysis, independent of the power-law assumption, for the first five

frequency bins of the NANOGrav 15-yr data set, which give the best constraints on the SGWB amplitude (Agazie et al. 2023b). All three of our models underpredict the amplitude of the SGWB (with respect to both the power-law and free-spectrum analyses) in the frequency range ($\sim 2 - 10$) nHz where NANOGrav is most sensitive.

This is consistent with previous predictions from cosmological simulations (see for example Kelley et al. 2016) and studies that suggest that BH mass relations need to have a significantly higher normalization, or significantly evolve with redshift, in order to match the PTA signal (Chen, Yu & Lu 2023b). Reducing the fixed hardening time-scale to 0.1 Gyr, as also studied by Agazie et al. (2023b), does not reconcile the predicted signal with the NANOGrav result. This highlights the necessity of using a more realistic hardening model for microphysical delays, opposed to the fixed hardening time used here, which can significantly impact the predicted amplitude of the SGWB.

We also note the possible impact that a limited box size may have on these predictions. As BHs at the massive end of the mass spectrum with $M_{\text{BH}} \gtrsim 10^{10} M_{\odot}$, which can have a significant impact on the amplitude of the SGWB, are missing from the cosmological box with side-length of $100 h^{-1}$ cMpc used for this study, due to their low number density. We have investigated the impact of adding such high-mass BHs to the FABLE population and found that their inclusion can amplify the predicted SGWB signal at the low end of the frequency range, but that this amplification depends on the details of the modelling choices made for this high-mass population. A detailed investigation of the effect of the limited box size on predictions of the SGWB is left for future work.

Eccentric orbits, expected at large binary separations but not modelled here, could boost the SGWB amplitude by shortening binary lifetimes and increasing merger rates (Kelley et al. 2017; Mannerkoski et al. 2022). On the other hand, if binaries remain eccentric up until the GW-dominated stage of the binary evolution, this leads to a suppression of the GW signal due to faster orbital decay, which would further increase the discrepancy with observations (Sesana 2013; Chen, Sesana & Del Pozzo 2017).

Focusing on the shape of the spectrum, all three of our models deviate from the expected $f^{-2/3}$ power law at the high-frequency end. This is expected, as at the high-frequency end the population is incomplete, with higher mass binaries merging before reaching this frequency (Sesana, Vecchio & Colacino 2008; Sato-Polito & Zaldarriaga 2025). At the low-frequency end, all three models slightly diverge from this simple power law due to environmental effects since the inspiral at large separations is not purely driven by GW radiation. However, this behaviour is highly sensitive to the chosen sub-grid hardening model, and can also be affected by eccentric orbits (Chen et al. 2017).

3.3.2 Implications for EM observations

The addition of macrophysical time delays not only impacts GW observations, as discussed earlier, but also influences key astrophysical properties of the SMBH population relevant for EM observations.

To illustrate this point, the distributions of Eddington fractions, as defined in Section 2.1.2, are shown in Fig. 14. These distributions show Eddington fractions of dual AGN, as we group together the Eddington fractions of both the primary and secondary BHs involved in a merger event. To obtain the distribution from the unmodified FABLE simulation, we use the instantaneous accretion rate of the BHs in the snapshot directly preceding the numerical BH merger. In our post-processed accretion model, we use the accretion rate calculated from gas cells in the central regions of the host subhaloes as

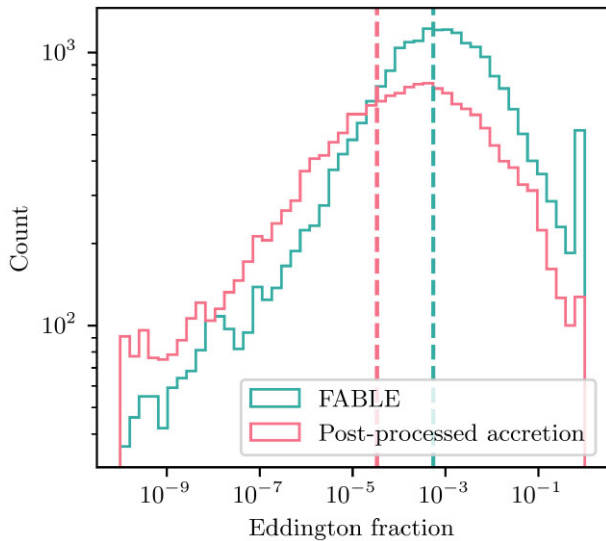


Figure 14. Distribution of Eddington fractions for dual AGN calculated in the snapshot before numerical BH merger for both the primary and secondary BHs in FABLE (teal line) compared to our model with macrophysical delays (pink line) where Eddington fractions are evaluated in the snapshot before host subhaloes merge. The vertical dashed lines show the median of the corresponding distribution.

described in Section 3.2.2, in the snapshot before the subhalo merger. For merger events whose host galaxies do not merge by $z = 0$, we take the accretion rate in the last snapshot of the simulation. Note that the Eddington ratios of these dual AGNs are measured when the BHs are at separations as shown in Fig. 8.

In our model, the median Eddington fraction of dual AGN decreases by around one order of magnitude compared to FABLE. This aligns with the observed redshift evolution of Eddington fractions. At high redshifts, the central regions of galaxies contain abundant cold gas that efficiently accretes onto BHs, whereas at lower redshifts, star formation and feedback processes suppress gas inflow, reducing accretion rates and lowering Eddington fractions (Shen & Kelly 2012). By delaying BH mergers, macrophysical time delays shift these events to lower redshifts, naturally leading to lower Eddington fractions. The difference in medians between the two distributions in Fig. 14 is comparable to the change in median Eddington fraction between $z = 1$ and 2 in FABLE, which corresponds approximately to the median macrophysical delay of 1.3 Gyr. Additionally, since our model incorporates macrophysical time delays on galactic scales, resulting in a higher number of binary SMBHs that do not merge by $z = 0$, we expect an increased prevalence of dual AGN at lower redshifts. In summary, our model predicts a population of dual AGN with lower luminosities, at lower redshifts and with larger separations than that inferred directly from FABLE.

Note that the peak at high Eddington fractions in the FABLE distribution (Fig. 14) arises because we are specifically considering BHs hosted in galaxies undergoing a merger event or close encounter. During galaxy mergers, strong tidal forces remove angular momentum from gas, driving it towards the central SMBHs and triggering AGN activity, which enhances accretion rates and increases Eddington fractions (Toomre & Toomre 1972; Barnes & Hernquist 1991, 1996; Talbot, Sijacki & Bourne 2024). In contrast, this peak is significantly less pronounced in our post-processed model, however it is not straightforward to isolate a single physical interpretation of this result. As highlighted in Fig. 10 and the corresponding discussion, more

gas infalls onto the secondary BH, since no BH particle is present in the original simulation and thus no AGN feedback is available to regulate gas inflow into the centre of the subhalo. However, the presence of an overmassive remnant BH in the primary halo may have the opposite effect on the accretion rate of the primary BH in our model, suppressing it through excessive feedback.³ This limitation underscores the challenges of implementing macrophysical time delays and accretion corrections in post-processing to account for premature BH mergers. In such models, AGN feedback cannot be consistently tracked, which in turn alters the accretion history of the constituent BHs together with key host galaxy properties.

4 DISCUSSION AND CONCLUSIONS

In this work, we use the FABLE cosmological boxes with periodic lengths of 40 and $100 h^{-1}$ cMpc (Henden et al. 2018, 2019, 2020; Bigwood et al. 2025) to study the population of merging SMBHs. We begin by examining the key properties of the BH population within the simulation and validating them against observations. Our analysis shows that FABLE reproduces the observed present-day BHMD and AGN luminosity function very well. However, it underpredicts the number density of low-mass BHs and overpredicts the number density of the most massive BHs compared to the observationally inferred BHMF. Additionally, FABLE reproduces the $M_{\text{BH}} - M_{\text{bulge}}$ relation well, but tends to underestimate BH masses for low-mass galaxies (i.e. $M_{\text{star, HM}} \lesssim 5 \times 10^{10} M_{\odot}$). These results put FABLE on par with other widely used cosmological simulations, with no simulation currently being able to accurately predict all observed SMBH population properties (Habouzit et al. 2021).

Focusing on the population of merging BHs, we have found that the majority of BHs merge prematurely in relation to the merger of their host galaxies in FABLE due to two compounding numerical effects.

(i) In the simulation, two BHs merge when they come within a smoothing length of each other, irrespective of whether they are gravitationally bound or not. This can occur while their host subhaloes are still interacting (e.g. at the first passage) but have not yet fully merged.

(ii) During subhalo interactions, strong variations in gravitational potentials, or steep potentials of nearby subhaloes can cause BHs to be gradually ‘repositioned’ to a subhalo other than their original host.

These numerical effects are also expected to play a role in other widely used cosmological simulations that implement a similar repositioning scheme as is used in FABLE. These include ILLUSTRIS, ILLUSTRISTNG, MILLENIUMTNG, EAGLE, FLAMINGO, and BAHAMAS, some of which have been used in previous works to make predictions of SMBH merger rates (Kelley et al. 2016; McCarthy et al. 2016; Salcido et al. 2016; DeGraf & Sijacki 2019; Siwek et al. 2020; Li et al. 2022; Pakmor et al. 2023; Schaye et al. 2023). We have conducted the first study that quantifies these numerical effects and provide an initial attempt at correcting for them. We do so by adding macrophysical time delays that correspond to the time elapsed in the simulation between numerical BH merger and the merger of the host subhaloes, which we track throughout the merger process. During these delays, BHs are allowed to grow in post-processing via Eddington limited,

³Even though these primary BHs may grow more since the accretion rate scales as M_{BH}^2 .

Bondi-like accretion based on the central gas properties of their host subhaloes. Our main findings are summarized below.

(i) These macrophysical time delays affect 93 per cent of the BH merging population we study, in which we consider all BH mergers in FABLE with $M_{\text{BH}} > 10^6 M_{\odot}$ at the time of numerical merger, residing in sufficiently massive subhaloes. These delays precede any microphysical time delays arising from unresolved small-scale hardening processes and are on average of the order of a few Gyrs, extending to $\gtrsim 12$ Gyr, a timespan over which host galaxies undergo significant morphological changes.

(ii) We argue that the BH separations at the time of numerical BH mergers suffer from the same numerical artefacts as the merger times themselves. Using these as initial separations for calculating subgrid hardening time-scales may lead to systematic errors. A more reliable, albeit conservative, alternative is to use the subhalo (galaxy) separations from the snapshot immediately before the subhaloes merge. These separations are, on average, two orders of magnitude larger than the BH separations, with clear implications for the length of microphysical time delays.

(iii) The introduction of macrophysical time delays suppresses high-mass merger events. Although accretion during the delay pushes surviving mergers toward higher chirp masses, it does not fully compensate for the suppression of the most massive events. Consequently, *LISA* will detect a higher percentage of SMBH mergers, leading to a more complete view of the merging SMBH population.

(iv) The reduction in BH merger rates, specifically in the highest mass bin, will directly affect the SGWB detected by PTAs, with a suppression observed across the full detectable frequency range. Our findings also illustrate how assumptions about SMBH hardening mechanisms shape the predicted SGWB signal.

(v) The large macrophysical time delays imply an increase in the number of observable dual AGN at later cosmic times. Moreover, these dual AGNs are predicted to be fainter than those in the original FABLE simulation, consistent with the observed redshift evolution of Eddington fractions, and residing at larger separations.

It is important to note that our method assumes that our substructure finding algorithm can accurately identify these substructures which host BH particles. Substructure finding algorithms may struggle to identify subhaloes in the central region of the host halo and rather associate the corresponding particles to the central subhalo (for a recent study on this, see Moreno et al. 2025). If the central subhalo and any of these neighbouring subhaloes host a BH, our method may assume an earlier host galaxy merger time if the SUBFIND algorithm is unable to resolve the two structures. Thus, using more accurate subhalo-finding algorithms might increase our macrophysical delays in certain cases. Furthermore, all substructure finding algorithms will struggle with subhalo identification for low-mass objects close to the resolution limit of simulations. Unfortunately, these are precisely the objects that host low-mass or seed BHs which recent *JWST* observations are starting to probe, indicating that very high resolution simulations are needed to understand this interesting BH mass regime.

In conclusion, our findings demonstrate that numerical choices in BH modelling within cosmological simulations significantly impact the predicted properties of merging SMBH populations. In this work, we quantify this issue and present a first attempt at mitigating for incorrect BH positions using a post-processing approach. However, the accuracy with which cosmological simulations capture the interplay between central BHs and their host galaxies cannot be fully reproduced through post-processing alone, underscoring a fundamental limitation of this method. Therefore, while these simulations remain a powerful tool for predicting current and future

GW and EM observations, our results highlight the pressing need for more precise modelling of both BH and BH binary orbits, both above and below the resolution limit of simulations.

ACKNOWLEDGEMENTS

The authors would like to thank Leah Bigwood for allowing us to use data from the FABLE $(100 h^{-1} \text{ cMpc})^3$ cosmological box in this work, and for all her valuable assistance in helping us understand this data. We also thank Stephen Taylor and Sophie Koudmani for valuable comments on the paper.

SB is supported by the Cambridge Centre for Doctoral Training in Data-Intensive Science funded by the UK Science and Technology Facilities Council (STFC). DS and MAB acknowledge support from the STFC, grant code ST/W000997/1. MAB additionally acknowledges support from a UKRI Stephen Hawking Fellowship (EP/X04257X/1). This work performed using the following facilities: the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service (www.csd3.cam.ac.uk), provided by Dell EMC and Intel using Tier2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/P020259/1), and DiRAC funding from the STFC (www.dirac.ac.uk), as well as the DiRAC@Durham facility operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure.

DATA AVAILABILITY

The FABLE data used in this paper are currently not publicly available through a web server, but is available upon request. All the scripts used to obtain the results in this work can be found at this [GitHub repository](#) and with (very minor) modification can be run on the ILLUSTRIS and ILLUSTRISTNG data which are already publicly available.

REFERENCES

- Abbott B. P. et al., 2016, *Phys. Rev. Lett.*, 116, 061102
- Abbott R. et al., 2020, *Phys. Rev. Lett.*, 125, 101102
- Abuter R. et al., 2024, *Nature*, 627, 281
- Agazie G. et al., 2023a, *ApJ*, 951, L8
- Agazie G. et al., 2023b, *ApJ*, 952, L37
- Amaro-Seoane P. et al., 2017, preprint (arXiv:1702.00786)
- Antoniadis J. et al., 2023, *A&A*, 678, A50
- Antonini F., Merritt D., 2012, *ApJ*, 745, 83
- Armitage P. J., Natarajan P., 2005, *ApJ*, 634, 921
- Artymowicz P., Lubow S. H., 1994, *ApJ*, 421, 651
- Athanassoula E., Fady E., Lambert J. C., Bosma A., 2000, *MNRAS*, 314, 475
- Bahé Y. M. et al., 2022, *MNRAS*, 516, 167
- Bardeen J. M., Petterson J. A., 1975, *ApJ*, 195, L65
- Barnes J. E., Hernquist L. E., 1991, *ApJ*, 370, L65
- Barnes J. E., Hernquist L., 1996, *ApJ*, 471, 115
- Begelman M. C., Blandford R. D., Rees M. J., 1980, *Nature*, 287, 307
- Bhowmick A. K. et al., 2024, *MNRAS*, 531, 4311
- Bigwood L., Bourne M. A., Irsic V., Amon A., Sijacki D., 2025, preprint (arXiv:2501.16983)
- Binney J., Tremaine S., 2011, *Galactic Dynamics: 2nd edn*. Princeton Series in Astrophysics. Princeton Univ. Press, Princeton, NJ, p. 647
- Bird S., Ni Y., Di Matteo T., Croft R., Feng Y., Chen N., 2022, *MNRAS*, 512, 3703

- Bogdan A. et al., 2024, *Nat. Astron.*, 8, 126
- Bourne M. A., Fiacconi D., Sijacki D., Piotrowska J. M., Koudmani S., 2024, *MNRAS*, 534, 3448
- Boylan-Kolchin M., Ma C.-P., Quataert E., 2007, *MNRAS*, 383, 93
- Callegari S., Kazantzidis S., Mayer L., Colpi M., Bellovary J. M., Quinn T., Wadsley J., 2011, *ApJ*, 729, 85
- Chabrier G., 2003, *PASP*, 115, 763
- Chandrasekhar S., 1943, *ApJ*, 97, 255
- Chen S., Sesana A., Del Pozzo W., 2017, *MNRAS*, 470, 1738
- Chen N., Ni Y., Tremmel M., Di Matteo T., Bird S., DeGraf C., Feng Y., 2022a, *MNRAS*, 510, 531
- Chen N. et al., 2022b, *MNRAS*, 514, 2220
- Chen N. et al., 2023a, *MNRAS*, 522, 1895
- Chen Y., Yu Q., Lu Y., 2023b, *ApJ*, 955, 132
- Chen N. et al., 2025, preprint (arXiv:2502.01024)
- Colpi M. et al., 2024, preprint (arXiv:2402.07571)
- Comerford J. M., Schluns K., Greene J. E., Cool R. J., 2013, *ApJ*, 777, 64
- Crain R. A. et al., 2015, *MNRAS*, 450, 1937
- Damiano A., Valentini M., Borgani S., Tornatore L., Murante G., Ragagnin A., Ragone-Figueroa C., Dolag K., 2024, *A&A*, 692, A81
- Davari R., Ho L. C., Peng C. Y., 2016, *ApJ*, 824, 112
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- DeGraf C., Sijacki D., 2019, *MNRAS*, 491, 4973
- Di Matteo T., Springel V., Hernquist L., 2005, *Nature*, 433, 604
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Dosopoulou F., Antonini F., 2017, *ApJ*, 840, 31
- Dubois Y., Devriendt J., Slyz A., Teyssier R., 2012, *MNRAS*, 420, 2662
- Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, *MNRAS*, 463, 3948
- EPTA Collaboration, InPTA Collaboration, 2024, *A&A*, 685, A94
- Euclid Collaboration, 2025, preprint (arXiv:2503.15317)
- Ferrarese L., 2002, *ApJ*, 578, 90
- Finn L. S., Thorne K. S., 2000, *Phys. Rev. D*, 62, 124021
- Franchini A., Lupi A., Sesana A., 2022, *ApJ*, 929, L13
- Gebhardt K. et al., 2000, *ApJ*, 539, L13
- Genel S. et al., 2014, *MNRAS*, 445, 175
- Genina A., Springel V., Rantala A., 2024, *MNRAS*, 534, 957
- Gerosa D., Veronesi B., Lodato G., Rosotti G., 2015, *MNRAS*, 451, 3941
- Gould A., Rix H.-W., 2000, *ApJ*, 532, L29
- Goulding A. D. et al., 2023, *ApJ*, 955, L24
- Gravity+ Collaboration, 2022, *The Messenger*, 189, 17
- Greene J. E., Strader J., Ho L. C., 2020, *ARA&A*, 58, 257
- Gültekin K. et al., 2009, *ApJ*, 698, 198
- Habouzit M. et al., 2021, *MNRAS*, 509, 3015
- Häring N., Rix H.-W., 2004, *ApJ*, 604, L89
- Henden N. A., Puchwein E., Shen S., Sijacki D., 2018, *MNRAS*, 479, 5385
- Henden N. A., Puchwein E., Sijacki D., 2019, *MNRAS*, 489, 2439
- Henden N. A., Puchwein E., Sijacki D., 2020, *MNRAS*, 498, 2114
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, *ApJS*, 175, 356
- Husa S., Khan S., Hannam M., Pürrer M., Ohme F., Forteza X. J., Bohé A., 2016, *Phys. Rev. D*, 93, 044006
- Inayoshi K., Ichikawa K., Ostriker J. P., Kuiper R., 2019, *MNRAS*, 486, 5377
- Inayoshi K., Visbal E., Haiman Z., 2020, *ARA&A*, 58, 27
- Juodžbalis I. et al., 2024, *Nature*, 636, 594
- Katz M. L., Larson S. L., 2018, *MNRAS*, 483, 3108
- Katz N., Weinberg D. H., Hernquist L., 1996, *ApJS*, 105, 19
- Kelley L. Z., Blecha L., Hernquist L., 2016, *MNRAS*, 464, 3131
- Kelley L. Z., Blecha L., Hernquist L., Sesana A., Taylor S. R., 2017, *MNRAS*, 471, 4508
- Khan F. M., Holley-Bockelmann K., Berczik P., Just A., 2013, *ApJ*, 773, 100
- Khan S., Husa S., Hannam M., Ohme F., Pürrer M., Forteza X. J., Bohé A., 2016a, *Phys. Rev. D*, 93, 044007
- Khan F. M., Fiacconi D., Mayer L., Berczik P., Just A., 2016b, *ApJ*, 828, 73
- King A. R., Kolb U., 1999, *MNRAS*, 305, 654
- Klein A. et al., 2016, *Phys. Rev. D*, 93, 024003
- Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
- Koudmani S., Henden N. A., Sijacki D., 2021, *MNRAS*, 503, 3568
- Koudmani S., Sijacki D., Smith M. C., 2022, *MNRAS*, 516, 2112
- Koudmani S., Somerville R. S., Sijacki D., Bourne M. A., Jiang Y.-F., Profit K., 2024, *MNRAS*, 532, 60
- Li K., Bogdanovic T., Ballantyne D. R., Bonetti M., 2022, *ApJ*, 933, 104
- Li K., Volonteri M., Dubois Y., Beckmann R., Trebitsch M., 2024, preprint (arXiv:2410.07856)
- Lodato G., Nayakshin S., King A. R., Pringle J. E., 2009, *MNRAS*, 398, 1392
- Lotz J. M., Jonsson P., Cox T. J., Primack J. R., 2008, *MNRAS*, 391, 1137
- Ludlow A. D., Schaye J., Schaller M., Richings J., 2019, *MNRAS*, 488, L123
- Ma L., Hopkins P. F., Kelley L. Z., Faucher-Giguère C.-A., 2023, *MNRAS*, 519, 5543
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Magorrian J. et al., 1998, *AJ*, 115, 2285
- Maiolino R. et al., 2024, *A&A*, 691, A145
- Mannerkoski M., Johansson P. H., Rantala A., Naab T., Liao S., 2021, *ApJ*, 912, L20
- Mannerkoski M., Johansson P. H., Rantala A., Naab T., Liao S., Rawlings A., 2022, *ApJ*, 929, 167
- Mannerkoski M., Rawlings A., Johansson P. H., Naab T., Rantala A., Springel V., Irodotou D., Liao S., 2023, *MNRAS*, 524, 4062
- Marconi A., Hunt L. K., 2003, *ApJ*, 589, L21
- Matteo T. D., Colberg J., Springel V., Hernquist L., Sijacki D., 2008, *ApJ*, 676, 33
- McCarthy I. G., Schaye J., Bird S., Le Brun A. M. C., 2016, *MNRAS*, 465, 2936
- McConnell N. J., Ma C.-P., 2013, *ApJ*, 764, 184
- Merloni A. et al., 2009, *ApJ*, 708, 137
- Merritt D., 1996, *AJ*, 111, 2462
- Merritt D., Milosavljević M., 2005, *Living Rev. Relativ.*, 8, 8
- Moreno V. J. F., Helly J., McGibbon R., Schaye J., Schaller M., Han J., Kugel R., 2025, preprint (arXiv:2502.06932)
- Nelson D. et al., 2019, *Comput. Astrophys. Cosmol.*, 6, 2
- Pacucci F., Nguyen B., Carniani S., Maiolino R., Fan X., 2023, *ApJ*, 957, L3
- Pakmor R., Springel V., Bauer A., Mocz P., Munoz D. J., Ohlmann S. T., Schaaf K., Zhu C., 2015, *MNRAS*, 455, 1134
- Pakmor R. et al., 2023, *MNRAS*, 524, 2539
- Perna M. et al., 2023, *A&A*, 679, A89
- Perna M. et al., 2025, *A&A*, 696, A59
- Peters P. C., 1964, *Phys. Rev.*, 136, B1224
- Petts J. A., Gualandris A., Read J. I., 2015, *MNRAS*, 454, 3778
- Petts J. A., Read J. I., Gualandris A., 2016, *MNRAS*, 463, 858
- Pfister H., Volonteri M., Dubois Y., Dotti M., Colpi M., 2019, *MNRAS*, 486, 101
- Phinney E. S., 2001, preprint (arXiv:astro-ph/0108028)
- Pillepich A. et al., 2017, *MNRAS*, 473, 4077
- Planck Collaboration XIII 2016, *A&A*, 594, A13
- Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, *MNRAS*, 338, 14
- Pringle J. E., 1991, *MNRAS*, 248, 754
- Puerto-Sánchez C. et al., 2025, *MNRAS*, 536, 3016
- Quinlan G. D., 1996, *New Astron.*, 1, 35
- Ragone-Figueroa C., Granato G. L., Ferraro M. E., Murante G., Biffi V., Borgani S., Planelles S., Rasia E., 2018, *MNRAS*, 479, 1125
- Rantala A., Pihajoki P., Johansson P. H., Naab T., Lahén N., Sawala T., 2017, *ApJ*, 840, 53
- Reardon D. J. et al., 2023, *ApJ*, 951, L6
- Reines A. E., Volonteri M., 2015, *ApJ*, 813, 82
- Richstone D. et al., 1998, *Nature*, 385, A14
- Rodriguez C., Taylor G. B., Zavala R. T., Peck A. B., Pollack L. K., Romani R. W., 2006, *ApJ*, 646, 49
- Rodriguez-Gomez V. et al., 2015, *MNRAS*, 449, 49
- Rosas-Guevara Y. M., Bower R. G., McAlpine S., Bonoli S., Tissera P. B., 2019, *MNRAS*, 483, 2712
- Ruiter A. J., Belczynski K., Benacquista M., Larson S. L., Williams G., 2010, *ApJ*, 717, 1006
- Ryan B. R., Ressler S. M., Dolence J. C., Tchekhovskoy A., Gammie C., Quataert E., 2017, *ApJ*, 844, L24
- Salcido J., Bower R. G., Theuns T., McAlpine S., Schaller M., Crain R. A., Schaye J., Regan J., 2016, *MNRAS*, 463, 870

- Salpeter E. E., 1955, *ApJ*, 121, 161
- Sato-Polito G., Zaldarriaga M., 2025, *Phys. Rev. D*, 111, 023043
- Sato-Polito G., Zaldarriaga M., Quataert E., 2024, *Phys. Rev. D*, 110, 063020
- Sayeb M., Blecha L., Kelley L. Z., Gerosa D., Kesden M., Thomas J., 2020, *MNRAS*, 501, 2531
- Scannapieco C., Gadotti D. A., Jonsson P., White S. D. M., 2010, *MNRAS*, 407, L41
- Schaye J. et al., 2023, *MNRAS*, 526, 4978
- Sesana A., 2013, *Class. Quantum Gravity*, 30, 224014
- Sesana A., Vecchio A., Colacino C. N., 2008, *MNRAS*, 390, 192
- Shankar F., 2013, *Class. Quantum Gravity*, 30, 244001
- 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
- Shen Y., Kelly B. C., 2012, *ApJ*, 746, 169
- Sijacki D., Springel V., Di Matteo T., Hernquist L., 2007, *MNRAS*, 380, 877
- Sijacki D., Springel V., Haehnelt M. G., 2009, *MNRAS*, 400, 100
- Sijacki D., Vogelsberger M., Genel S., Springel V., Torrey P., Snyder G. F., Nelson D., Hernquist L., 2015, *MNRAS*, 452, 575
- Siwek M. S., Kelley L. Z., Hernquist L., 2020, *MNRAS*, 498, 537
- Siwek M., Weinberger R., Hernquist L., 2023, *MNRAS*, 522, 2707
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, *MNRAS*, 391, 481
- Springel V., 2010, *MNRAS*, 401, 791
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Springel V., Di Matteo T., Hernquist L., 2005a, *MNRAS*, 361, 776
- Springel V. et al., 2005b, *Nature*, 435, 629
- Talbot R. Y., Sijacki D., Bourne M. A., 2024, *MNRAS*, 528, 5432
- Terrazas B. A. et al., 2020, *MNRAS*, 493, 1888
- Toomre A., Toomre J., 1972, *ApJ*, 178, 623
- Torrey P., Vogelsberger M., Genel S., Sijacki D., Springel V., Hernquist L., 2014, *MNRAS*, 438, 1985
- Tremaine S. et al., 2002, *ApJ*, 574, 740
- Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, *MNRAS*, 451, 1868
- Trindade Falcão A., Turner T. J., Kraemer S. B., Reeves J., Braito V., Schmitt H. R., Feuillet L., 2024, *ApJ*, 972, 185
- Übler H. et al., 2024, *MNRAS*, 531, 355
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031
- Vogelsberger M. et al., 2014, *MNRAS*, 444, 1518
- Volonteri M., Pfister H., Beckmann R., Dotti M., Dubois Y., Massonneau W., Musoke G., Tremmel M., 2022, *MNRAS*, 514, 640
- Weinberger R. et al., 2016, *MNRAS*, 465, 3291
- Wiersma R. P. C., Schaye J., Smith B. D., 2009a, *MNRAS*, 393, 99
- Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009b, *MNRAS*, 399, 574
- Xie F.-G., Yuan F., 2012, *MNRAS*, 427, 1580
- Xu H. et al., 2023, *Res. Astron. Astrophys.*, 23, 075024
- Yu Q., Tremaine S., 2002, *MNRAS*, 335, 965

APPENDIX A: REDSHIFT EVOLUTION OF THE BH MASS–STELLAR MASS SCALING RELATION

In Fig. A1, we show the redshift evolution of the BH mass relation, where we plot the BH mass as a function of the stellar mass within twice the stellar half-mass radius of the host galaxy. We do this for easier comparison to observations, rather than using a proxy for the bulge mass as was done in Fig. 4, as it is increasingly more challenging to separate the mass of the bulge from the total stellar mass of galaxies at high redshifts (Davari, Ho & Peng 2016). We compare these to observations by Reines & Volonteri (2015) and Greene et al. (2020) at $z = 0$. The cosmic evolution of this relation is not yet well understood, with recent *JWST* measurements portraying a complex picture (Maiolino et al. 2024), and thus it is difficult to compare the population of BHs in FABLE to the observed population of SMBHs in this respect (Merloni et al. 2009). However, this figure indicates that host galaxies in the simulation are growing more rapidly than their central BHs, especially at the low-mass end. Future observations with *JWST* and GRAVITY+ will provide better constraints on this mass relation up to $z \gtrsim 2$, improving our understanding of BH and galaxy co-evolution in the early universe (Gravity+ Collaboration 2022; Abuter et al. 2024).

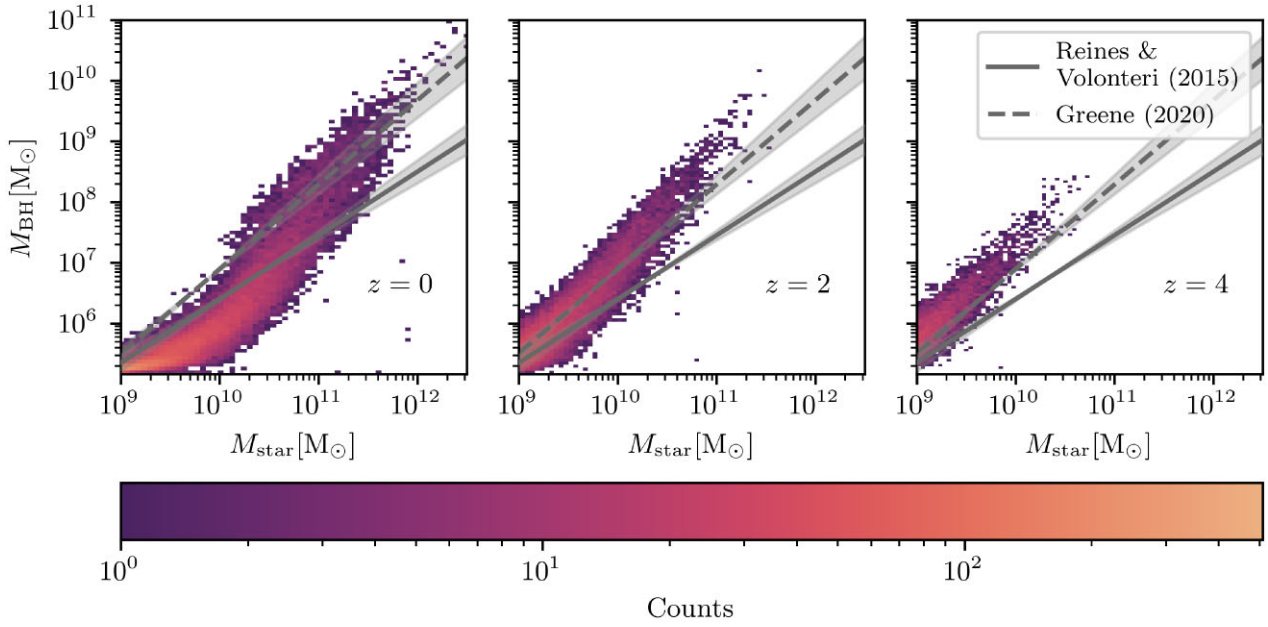


Figure A1. The evolution of the central BH mass versus stellar mass within twice the half-mass radius of all galaxies in FABLE. The redshifts shown are $z = 0$, 2, and 4. The best fits from Reines & Volonteri (2015) and Greene et al. (2020) at $z = 0$ are also shown for comparison.

This paper has been typeset from a \LaTeX file prepared by the author.