



# The Cosmic Evolution of the Supermassive Black Hole Population: A Hybrid Observed Accretion and Simulated Mergers Approach

Fan Zou<sup>1,2,3</sup> , W. N. Brandt<sup>2,3,4</sup> , Elena Gallo<sup>1</sup> , Bin Luo<sup>5,6</sup> , Qingling Ni<sup>7</sup> , Yongquan Xue<sup>8,9</sup> , and Zhibo Yu<sup>2,3</sup>

<sup>1</sup> Department of Astronomy, University of Michigan, 1085 S University, Ann Arbor, MI 48109, USA; [fanzou01@gmail.com](mailto:fanzou01@gmail.com)

<sup>2</sup> Department of Astronomy and Astrophysics, 525 Davey Lab, The Pennsylvania State University, University Park, PA 16802, USA

<sup>3</sup> Institute for Gravitation and the Cosmos, The Pennsylvania State University, University Park, PA 16802, USA

<sup>4</sup> Department of Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

<sup>5</sup> School of Astronomy and Space Science, Nanjing University, Nanjing, Jiangsu 210093, People's Republic of China

<sup>6</sup> Key Laboratory of Modern Astronomy and Astrophysics (Nanjing University), Ministry of Education, Nanjing 210093, People's Republic of China

<sup>7</sup> Max-Planck-Institut für extraterrestrische Physik (MPE), Gießenbachstraße 1, D-85748, Garching bei München, Germany

<sup>8</sup> CAS Key Laboratory for Research in Galaxies and Cosmology, Department of Astronomy, University of Science and Technology of China, Hefei 230026, People's Republic of China

<sup>9</sup> School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026, People's Republic of China

Received 2024 June 27; revised 2024 September 10; accepted 2024 September 23; published 2024 November 11

## Abstract

Supermassive black holes (SMBHs) can grow through both accretion and mergers. It is still unclear how SMBHs evolve under these two channels from high redshifts to the SMBH population we observe in the local Universe. Observations can directly constrain the accretion channel but cannot effectively constrain mergers yet, while cosmological simulations provide galaxy merger information but can hardly return accretion properties consistent with observations. In this work, we combine the observed accretion channel and the simulated merger channel, taking advantage of observations and cosmological simulations, to depict a realistic evolution pattern of the SMBH population. With this methodology, we can derive the scaling relation between the black hole mass ( $M_{\text{BH}}$ ) and host-galaxy stellar mass ( $M_*$ ), and the local black hole mass function (BHMF). Our scaling relation is lower than those based on dynamically measured  $M_{\text{BH}}$ , supporting the claim that dynamically measured SMBH samples may be biased. We show that the scaling relation has little redshift evolution. The BHMF steadily increases from  $z = 4$  to  $z = 1$  and remains largely unchanged from  $z = 1$  to  $z = 0$ . The overall SMBH growth is generally dominated by the accretion channel, with possible exceptions at high mass ( $M_{\text{BH}} \gtrsim 10^8 M_\odot$  or  $M_* \gtrsim 10^{11} M_\odot$ ) and low redshift ( $z \lesssim 1$ ). We also predict that around 25% of the total SMBH mass budget in the local Universe may be locked within long-lived, wandering SMBHs, and the wandering mass fraction and wandering SMBH counts increase with  $M_*$ .

*Unified Astronomy Thesaurus concepts:* Supermassive black holes (1663); Cosmological evolution (336)

## 1. Introduction

Supermassive black holes (SMBHs) are thought to reside in the centers of almost all massive galaxies and have significantly shaped the evolution of the Universe. In a cosmological context, SMBHs can grow through two channels—accretion and mergers. The accretion-driven growth primarily happens during the rapid-accretion phase when SMBHs are observed as active galactic nuclei (AGNs), and the merger-driven growth occurs after mergers of galaxies including their dark-matter halos, as predicted by the hierarchical galaxy formation model (e.g., S. D. M. White & M. J. Rees 1978). SMBHs follow these two growth channels to evolve from the early Universe to those observed in the local Universe. It is generally thought that SMBH growth is dominated by the accretion channel (e.g., A. Sotan 1982; A. Marconi et al. 2004; A. Küller et al. 2015; F. Pacucci & A. Loeb 2020) at most redshifts at which our current observational data are sensitive ( $z \lesssim 4$ ; e.g., F. Zou et al. 2024), beyond which we can generally only capture the most-luminous tip of the SMBH population.<sup>10</sup>

However, the merger channel may become important at the high-mass/low- $z$  regime, where the available accreting material may not be sufficient to support strong accretion-driven growth (e.g., D. Liu et al. 2019). Additionally, in the very early Universe when SMBH seeds were formed, they may also undergo an epoch when the merger channel dominates if their number density is high (e.g., F. Pacucci & A. Loeb 2020; A. Sicilia et al. 2022), but the prediction of this phase strongly depends upon the high-redshift SMBH seeding model (e.g., M. A. Latif & A. Ferrara 2016).

The SMBH population, although found six decades ago with the discovery of quasars (M. Schmidt 1963), still has many unclear properties. What is their number density as a function of the black hole mass ( $M_{\text{BH}}$ ) and redshift, i.e., the black hole mass function (BHMF)? How does the scaling relation between  $M_{\text{BH}}$  and host-galaxy stellar masses ( $M_*$ ) evolve with redshift? How large are the contributions of the two growth channels to the overall SMBH population growth? We aim to address these questions in this work; more broadly, we seek to understand how the SMBH population evolves across cosmic time. This can be done by tracing the SMBH growth and integrating from high redshift to the local Universe (e.g., A. Sotan 1982; A. Marconi et al. 2004), and the accretion and merger growth channels are discussed separately below.

The accretion channel can be directly measured from observations. To provide a broad idea, the SMBH accretion power can be traced by the AGN luminosity, which can be

<sup>10</sup> The James Webb Space Telescope (JWST) is capable of probing the SMBH growth at higher redshifts (e.g., R. Maiolino et al. 2023; G. Yang et al. 2023), but such campaigns have been advancing for only two years, and more work needs to be done before the community reaches a consensus.



converted to the black hole accretion rate (i.e.,  $dM_{\text{BH}}/dt$ ). When averaging over a large galaxy sample, the long-term population mean black hole accretion rate, denoted as  $\dot{M}_a$ , can be measured.  $\dot{M}_a$  has been shown to be closely correlated with  $M_*$  and redshift (e.g., G. Yang et al. 2017, 2018b; F. Zou et al. 2024), and thus we can use  $(M_*, z)$  as predictors of  $\dot{M}_a$  for the galaxy population. This is important because  $M_*$  and  $z$  are the most straightforward physical parameters to measure with modern multi-wavelength photometric surveys (e.g., F. Zou et al. 2022). In F. Zou et al. (2024), we provided the best current measurement of  $\dot{M}_a$  as a function of  $(M_*, z)$  based on a large compilation of data from nine high-quality survey fields (spanning  $0.05\text{--}60\text{ deg}^2$  following a wedding-cake design) and a semiparametric Bayesian method. F. Zou et al. (2024) adopted the X-ray emission as the accretion-power tracer, which is universal among AGNs, has high penetrating power through obscuration, and suffers little from starlight dilution (e.g., W. N. Brandt & D. M. Alexander 2015; Y. Q. Xue 2017; W. N. Brandt & G. Yang 2022). All of these factors enabled reliable measurements of  $\dot{M}_a$  across a wide parameter space, covering  $9.5 < \log(M_*/M_\odot) < 12$  and redshifts up to 4. F. Zou et al. (2024) also showed that this  $\dot{M}_a(M_*, z)$  function, after convolution with the galaxy stellar mass function (SMF), can successfully return an AGN X-ray luminosity function consistent with direct observations.

The merger channel, however, can hardly be directly constrained by observations, at least for now. In principle, SMBH mergers can be probed through nano-hertz gravitational waves (GWs) observed by pulsar timing arrays (e.g., see C. M. F. Mingarelli 2019 for a brief review), and recent results have indeed begun to provide promising insights (e.g., G. Agazie et al. 2023). However, the currently observed nano-hertz GW background still has a low signal-to-noise ratio and cannot provide effective constraints on SMBH mergers. Fortunately, the merger channel is accessible via cosmological simulations. Modern cosmological simulations are now able to self-consistently and simultaneously reveal both large structures of up to hundreds of megaparsecs and internal physics within individual galaxies. Galaxies merge during the simulated cosmic evolution, which may lead to SMBH mergers, but the eventual SMBH coalescence status depends upon whether there is sufficient time for the SMBHs to dissipate their orbital energy. Unlike accretion, which is directly linked to the vicinity of SMBHs, the overall galaxy merger history is largely set by the much larger environment at the dark-matter halo scale. Therefore, merger histories are generally more reliable in simulations compared to simulated accretion. We rely on the IllustrisTNG project (R. Weinberger et al. 2017; F. Marinacci et al. 2018; J. P. Naiman et al. 2018; D. Nelson et al. 2018; A. Pillepich et al. 2018a, 2018b; V. Springel et al. 2018; D. Nelson et al. 2019a, 2019b; A. Pillepich et al. 2019), a series of gravo-magnetohydrodynamical cosmological simulations, to extract the merger information.<sup>11</sup>

In this work, we use a hybrid method to combine the accretion channel from observations and the merger channel from simulations, as discussed above, to integrate over time and predict the overall cosmic evolution of the SMBH population. There have been valuable efforts to understand the SMBH population evolution from either the observational or the theoretical sides, although the exact boundaries between theories and observations are sometimes vague. Indeed, the SMBH population is widely

implemented in semi-analytic models (SAMs) and hydrodynamical simulations for galaxy evolution (e.g., A. Kulier et al. 2015; D. Sijacki et al. 2015; R. S. Somerville & R. Davé 2015; R. Weinberger et al. 2017; A. Ricarte & P. Natarajan 2018). However, these theoretical works inevitably have to simplify the complex physics governing the SMBH and galaxy evolution, which may cause challenges in having realistic characterization of SMBHs. For example, IllustrisTNG and many other cosmological simulations have already included the SMBH population, which means that, in principle, they can model the SMBH accretion. However, given the limited resolution, these simulated SMBHs are primarily modeled with largely uncertain sub-grid physics. It should also be noted that these large-scale cosmological simulations are not calibrated against the observed AGN population such as the AGN luminosity function or  $\dot{M}_a(M_*, z)$ . Even for detailed, resolved modeling focusing on the AGN disk and feedback, simulations still have significant inconsistencies compared with observations (e.g., A. Lawrence 2018; S. W. Davis & A. Tchekhovskoy 2020; and references therein). It has been shown that the SMBH population in cosmological simulations cannot match well with observations (e.g., J. Aird et al. 2018; M. Habouzit et al. 2021, 2022), and we will also illustrate this point in our following sections. As a complementary approach, observational constraints generally do not require assumptions about the underlying physics, but instead connect the  $M_{\text{BH}}$  growth to accretion-driven electromagnetic radiation or host-galaxy properties (e.g., A. Sotan 1982; A. Marconi et al. 2004; F. Shankar et al. 2009; G. Yang et al. 2018b). Some models also combine observations and theories to characterize SMBHs. For example, A. Ricarte & P. Natarajan (2018) implemented the observation-motivated accretion channel in their SAM, F. Shankar et al. (2020b) combined the observed accretion channel and the SAM-predicted galaxy evolutionary tracks, and H. Zhang et al. (2023) presented a data-driven empirical model constrained by various observed results. These approaches generally show more realistic modeling of the SMBH and AGN population than purely theoretical results. However, such observational constraints usually lack direct, careful treatments of mergers. Due to these reasons, we aim to take advantage of both observations and simulations to have a more realistic depiction of SMBH growth. Instead of focusing on setting appropriate models for the SMBH accretion to match observations, we will directly implement the observed accretion channel; when combined with simulations, we can further derive the contribution from the merger channel. Besides, this work benefits from both the state-of-the-art observational and simulation results that became available only within recent years. Our adopted new results in F. Zou et al. (2024) significantly improve the constraints for the accretion channel, and the IllustrisTNG simulations have been intensively explored and proven to be successful over the past six years.

This work is structured as follows. Section 2 describes the methodology. Section 3 presents our results and discusses their implications. Section 4 summarizes this work. We adopt a flat  $\Lambda$ CDM cosmology with  $H_0 = 70\text{ km s}^{-1}\text{ Mpc}^{-1}$ ,  $\Omega_\Lambda = 0.7$ , and  $\Omega_M = 0.3$ .

## 2. Methodology

### 2.1. Overview

Briefly, to reconstruct the SMBH population at different redshifts, we begin at  $z = 4$  with some given initial conditions and let SMBHs grow through both the accretion and merger channels to evolve to  $z = 0$ . We only probe redshift up to 4

<sup>11</sup> <https://www.tng-project.org>

because we do not have reliable observational constraints at higher redshift (e.g., F. Zou et al. 2024).

The TNG simulations provide properties including  $M_*$  of each galaxy as a function of  $z$  and their merger trees. Starting at  $z=4$ , we seed SMBHs<sup>12</sup> into the TNG-simulated galaxies that have  $M_* > 10^{9.5} M_\odot$  for the first time using the  $M_{\text{BH}}-M_*$  scaling relation calibrated against broad-line AGNs in A. E. Reines & M. Volonteri (2015),  $\log(M_{\text{BH}}/M_\odot) = 7.45 + 1.05[\log(M_*/M_\odot) - 11]$ , as will be discussed in Section 2.2. We add a typical scatter of 0.5 dex to the seeded  $M_{\text{BH}}$  (e.g., A. E. Reines & M. Volonteri 2015; J. I. H. Li et al. 2023). This means that, at  $z=4$ , we seed all the galaxies with  $M_*(z=4) > 10^{9.5} M_\odot$  while setting the  $M_{\text{BH}}$  of the remaining less-massive galaxies to be 0; once a low-mass galaxy accumulates sufficient  $M_*$  to reach  $10^{9.5} M_\odot$  at lower redshift, we then seed an SMBH into it. Therefore, considering the main progenitors, SMBHs are not necessarily seeded at  $z=4$ . Section 3.4 will present typical  $M_*$  evolutionary tracks. Briefly, galaxies with  $M_*(z=0) > 10^{10.5} M_\odot$  generally reach  $10^{9.5} M_\odot$  at  $z \gtrsim 3$  and are thus seeded sufficiently early such that the majority of their SMBH growth is captured. However, low-mass galaxies with  $M_*(z=0) < 10^{10} M_\odot$  reach  $10^{9.5} M_\odot$  much later ( $z \lesssim 1$ ), and hence a considerable fraction of their SMBH growth may be missed in our analyzes, leading our results in the low-mass regime to be more dominated by the choice of seeding.

We let the SMBHs grow with a rate predicted by the median  $\dot{M}_a(M_*, z)$  function in Section 3.3 of F. Zou et al. (2024). When two galaxies merge according to the TNG merger trees, we either let the two SMBHs (if any) in these galaxies merge or only keep the more massive one in the newly merged galaxy, as will be detailed in Section 2.4. For a given galaxy labeled with a subscript of  $i$  evolving from  $z$  to  $z - \delta z$ , the above procedures can be summarized as the following equation,

$$M_{\text{BH},i}(z - \delta z) = M_{\text{BH},i}(z) + \dot{M}_a(M_{*,i}(z), z) \delta t + M_{\text{BH},i}^{\text{merger}}(z \rightarrow z - \delta z), \quad (1)$$

where  $\delta t$  is the cosmic time corresponding to  $\delta z$ , and  $M_{\text{BH},i}^{\text{merger}}(z \rightarrow z - \delta z)$  is the increase in  $M_{\text{BH},i}$  from  $z$  to  $z - \delta z$  through mergers. There are 79 TNG snapshots at  $z \leq 4$  with a time step of  $\approx 0.15$  Gyr. We iteratively follow these procedures until  $z=0$  to recover the expected SMBH population in the range  $0 \leq z \leq 4$ . This approach of integrating the accretion across time is broadly related to the classical Soltan argument (A. Soltan 1982), which has been written in differential forms (e.g., T. A. Small & R. D. Blandford 1992; A. Marconi et al. 2004) and also modified in F. Shankar et al. (2020b) to integrate along  $M_*$  evolutionary tracks. There are several important points worth noting regarding these procedures, and we will discuss them in the subsequent subsections.

## 2.2. SMBH Seeds

We only seed SMBHs into galaxies more massive than  $10^{9.5} M_\odot$ , which means that we do not attempt to probe less-massive galaxies. This is because the massive black holes (MBHs)<sup>13</sup> in dwarf galaxies with  $M_* < 10^{9.5} M_\odot$  are still poorly

understood, and we currently do not have a well-measured  $\dot{M}_a(M_*, z)$  function in this low-mass regime (e.g., F. Zou et al. 2023, 2024). In fact, dwarf galaxies do not always necessarily host MBHs, and observational constraints on their MBH-occupation fraction are still loose (e.g., B. P. Miller et al. 2015; E. Gallo & A. Sesana 2019). The accretion-driven growth of MBHs in dwarf galaxies also does not necessarily mainly happen through typical AGN gas accretion but may be substantially powered by tidal-disruption events (e.g., K. Zubovas 2019).

Our seeding follows the  $M_{\text{BH}}-M_*$  relation calibrated against broad-line AGNs in A. E. Reines & M. Volonteri (2015). They reported separate scaling relations for AGNs and inactive galaxies with dynamical  $M_{\text{BH}}$  measurements and showed that these two relations can differ by over 1 order of magnitude. F. Shankar et al. (2016, 2020b, 2020a) argued that the AGN one is more accurate because the relation based on dynamically measured  $M_{\text{BH}}$  may be significantly biased toward those with more massive SMBHs whose spheres of influence are easier to resolve. Our results later (Section 3.1) also seem to be more consistent with the AGN-based relation. We will discuss the impacts of the adopted seeding relation on our results in Section 3.7.

Furthermore, by using the local-Universe  $M_{\text{BH}}-M_*$  relation at all redshifts, we implicitly adopt the assumption that this scaling relation does not evolve with redshift, at least up to  $z=4$ . This is a strong assumption that needs careful assessment. It has been shown in recent works that this scaling relation does not show a significant redshift evolution (e.g., M. Sun et al. 2015; G. Yang et al. 2018b, 2019; F. Shankar et al. 2020b; H. Suh et al. 2020; J. Li et al. 2021a; J. I. H. Li et al. 2021b; M. Habouzit et al. 2021; G. Yang et al. 2022; J. I. H. Li et al. 2023), though direct observational assessments can hardly reach  $z > 1$ . Our results in Section 3.1 will show that our predicted scaling relation indeed does not have a large variation across different redshifts. One side point worth noting is that, at higher redshifts above  $z=4$ , there is some emerging evidence suggesting that SMBHs may be over-massive (e.g., F. Pacucci & A. Loeb 2024), but it is still under debate whether the high-redshift  $M_{\text{BH}}-M_*$  relation is indeed higher than the local-Universe one, and the corresponding possible evidence is mainly from the early Universe, which is beyond the scope of this work and would not materially affect our conclusions. We will also try a redshift-dependent seeding relation in Section 3.7.

Furthermore, we adopt a single typical seeding scatter of 0.5 dex for all the redshifts. It is unclear how large the  $M_{\text{BH}}-M_*$  scatter is at higher redshifts, and different simulations return different results (M. Habouzit et al. 2021). Nevertheless, we expect the seeding scatter to contribute little to our overall  $M_{\text{BH}}$  scatter at  $M_* \gtrsim 10^{10} M_\odot$  and  $z < 3$ , as explored in Section 3.1.

## 2.3. The Accretion Channel

Our chosen growth rate from Section 3.3 of F. Zou et al. (2024) only depends on  $(M_*, z)$  but not on the SMBH itself or its vicinity. This approach has an important advantage in that  $M_*$ , as a global parameter of a galaxy, is generally more reliable to measure than local parameters, and utilizing the observational constraints in F. Zou et al. (2024) avoids relying on oftentimes largely uncertain sub-grid physics to depict the SMBH accretion, as done in simulations.

<sup>12</sup> Note that the term seed in our context is irrelevant to the earliest SMBH seeds at even higher redshift. We only mean that we set  $M_{\text{BH}}$  for massive galaxies as initial conditions.

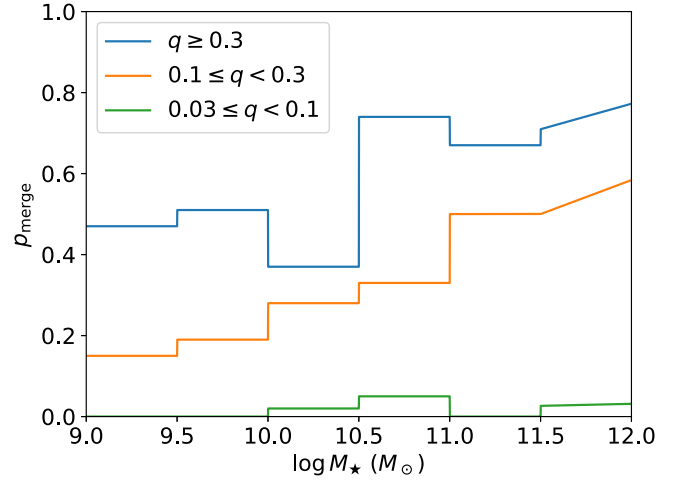
<sup>13</sup> The term SMBH is usually not used for these dwarf galaxies because their hosted black holes may not be sufficiently massive to be SMBHs but may instead be intermediate-mass black holes. We thus use the general term MBH here.



The  $\dot{M}_a$  in F. Zou et al. (2024) is the mean  $M_{\text{BH}}$  growth rate averaged over all the galaxies with given  $(M_*, z)$ . The short-term AGN phase has been averaged out during this process, and thus  $\dot{M}_a$  reflects the long-term population mean. X-rays were used to sample the accretion power, and there are two additional quantities assumed. One is the bolometric correction converting X-ray luminosities to bolometric luminosities, and the luminosity-dependent bolometric correction in F. Duras et al. (2020) was adopted. The other is the accretion radiative efficiency converting bolometric luminosities to mass-accretion rates, which is assumed to be a canonical value of 0.1 (e.g., S. W. Davis & A. Laor 2011). These conversions have typical systematic uncertainties up to a factor of a few. An additional systematic uncertainty arises from the fact that the X-ray-constrained  $\dot{M}_a$  in F. Zou et al. (2024) may miss Compton-thick accretion. The typical fraction of Compton-thick AGNs at  $z = 0$  is generally constrained to be  $\approx 20\% - 50\%$  (e.g., C. Ricci et al. 2015; E. S. Kammoun et al. 2020; Boorman et al., submitted). Using the X-ray luminosity function with the column-density distribution included in Sections 3 and 6 of Y. Ueda et al. (2014), we found that the fractional accretion power at column densities above  $10^{24} \text{ cm}^{-2}$  is  $\approx 38\%$  for all the redshifts. Besides, some Compton-thick AGNs are still able to be detected (e.g., J. Li et al. 2019; W. Yan et al. 2023), especially given the fact that we are probing higher rest-frame energies (reaching 15–40 keV) with higher penetrating power as redshift increases. Therefore, the systematic bias of  $\dot{M}_a$  from possibly missed Compton-thick accretion should be  $\lesssim 0.2$  dex, where the upper limit, 0.2 dex, corresponds to the case where all the Compton-thick accretion accounting for 38% of the total accretion power is missed. One caveat worth noting is that there is emerging evidence suggesting that more accretion power may be hidden by heavy obscuration at high redshift than our previous expectations. For example, G. Yang et al. (2021) showed that the X-ray-inferred  $\dot{M}_a$  may be underestimated by a factor of a few at  $z > 3$ , and a large number of little red dots at  $z \gtrsim 4$  were recently found by JWST and might contain heavily obscured AGNs (e.g., D. D. Kocevski et al. 2024). However, this possible discrepancy seems to mainly exist at high redshift and barely impacts the redshift range on which we are focusing, where most cosmic SMBH growth occurs.

Besides the dependence on  $(M_*, z)$ ,  $\dot{M}_a$  also depends on star formation rate (e.g., G. Yang et al. 2019), galaxy morphology (e.g., Q. Ni et al. 2021), and perhaps cosmic environment (e.g., H. Song et al. 2016; G. Yang et al. 2018a). However, for the general galaxy population, the dependence on star formation rate and environment is only secondary compared to the  $M_*$  dependence (e.g., G. Yang et al. 2017, 2018a; A. Amiri et al. 2019; G. Yang et al. 2019), and the morphological dependence, though possibly more fundamental, is still not fully understood in a quantitative sense because morphological measurements are expensive to obtain. Therefore,  $(M_*, z)$  should still be the best available predictors for  $\dot{M}_a$  that capture most of the variance.

There are statistical uncertainties when measuring  $\dot{M}_a(M_*, z)$  in F. Zou et al. (2024), which are not included in our analyses. Appendix A shows that these would only add a scatter of  $\approx 0.05 - 0.1$  dex into our  $M_{\text{BH}}$  and are thus negligible compared to the aforementioned systematic uncertainties.



**Figure 1.** Our adopted  $p_{\text{merge}}$  based on M. Tremmel et al. (2018b) as a function of  $M_*$  at different  $q$  bins. Term  $p_{\text{merge}}$  is set to 0 for  $q < 0.03$  and is extrapolated only in the massive end ( $M_* > 10^{11.5} M_\odot$ ).

#### 2.4. SMBH Mergers

Regarding the SMBHs after two galaxies merge, they usually undergo a long process before they finally reach coalescence, as discussed in, for example, Q. Yu (2002), E. Barausse & A. Lapi (2021), and references therein. The two SMBHs will first wander independently in the merged galaxy with a 10 kpc scale separation and gradually lose their orbital energy through the Chandrasekhar dynamical friction before sinking into the galaxy center and forming a close SMBH pair with a subkiloparsec separation (e.g., F. Dosopoulou & F. Antonini 2017). The SMBHs then harden at a parsec-scale separation where the dynamical friction process becomes inefficient, and the orbital energy will be dissipated through dynamical interactions with stars (e.g., G. D. Quinlan 1996), gas (e.g., G. Lodato et al. 2009), or even another SMBH (e.g., M. Bonetti et al. 2019). Once the binary separation shrinks to  $\approx 10^{-3} - 10^{-2}$  pc, the orbit will rapidly decay through GW emission and finally lead to an SMBH merger event. The whole process may take longer than the Hubble time, in which case the SMBH merger will be stalled (e.g., F. Dosopoulou & F. Antonini 2017; M. Tremmel et al. 2018b; A. Ricarte et al. 2021a, 2021b), and thus SMBHs only have a less-than-unity probability of eventually merging ( $p_{\text{merge}}$ ) after their host galaxies merge. Those that do not merge will constitute a long-lived wandering SMBH population.

We adopt  $p_{\text{merge}}$  in Figure 1 of M. Tremmel et al. (2018b), who presented this quantity as a function of the  $M_*$  ratio  $q$  ( $\leq 1$ ) of the two merging galaxies and the  $M_*$  of the primary galaxy within  $q > 0.03$  and  $10^9 < M_* < 10^{11.5} M_\odot$ . Such an approach was also adopted in F. Pacucci & A. Loeb (2020). The  $p_{\text{merge}}$  relation in M. Tremmel et al. (2018b) is derived from simulations with emphasis on modeling SMBHs such as their dynamics in the context of cosmological simulations and should be one of the best ones currently available. Term  $p_{\text{merge}}$  strongly increases with  $q$  and  $M_*$ , where  $p_{\text{merge}} \approx 0$  for minor mergers with  $0.03 < q < 0.1$  and reaches  $\approx 0.7$  for major mergers in massive galaxies with  $M_* > 10^{10.5} M_\odot$ . Therefore, we set  $p_{\text{merge}} = 0$  for  $q < 0.03$ , which is not covered in M. Tremmel et al. (2018b); we also use a linear relation to fit the  $p_{\text{merge}} - \log M_*$  relation at a given  $q$  bin in M. Tremmel et al. (2018b) and extrapolate it into  $M_* > 10^{11.5} M_\odot$ . We show  $p_{\text{merge}}$  versus  $M_*$  in Figure 1. In principle,  $p_{\text{merge}}$  also strongly

depends upon the galaxy structure such as the central stellar density and spherical asymmetry (e.g., E. Vasiliev et al. 2014; M. Tremmel et al. 2018b), but it is difficult to directly apply these relations, and we reiterate that our analyzes only aim to work in an average sense. Besides, M. Tremmel et al. (2018b) only accounted for the SMBH-binary evolution down to a separation of  $\approx 700$  pc. This indicates that the resulting  $p_{\text{merge}}$  is only an upper limit, although the SMBH-binary evolution time is generally expected to be relatively shorter than the evolution time before reaching this separation. We note that the evolution of SMBH binaries from a galactic scale down to a final coalescence is still under active research and beyond the scope of this work, and the merger rate heavily depends on the complex physics briefly outlined above, as extensively explored in, for example, E. Barausse et al. (2020). It often cannot be fully modeled even at a galactic scale in large cosmological simulations; for example, the TNG simulations themselves adopt a simple, so-called *repositioning* strategy to manually place the SMBH positions into the minimum of the gravitational potential at each time step (R. Weinberger et al. 2017), which would cause a much overestimated SMBH merger rate (N. Chen et al. 2022). Here, we turn to reasonable simplified approximations to quantify  $p_{\text{merge}}$  and will also explore the sensitivity of our results on the choice of  $p_{\text{merge}}$ . Especially, we will try extreme cases of  $p_{\text{merge}} = 0$  and  $p_{\text{merge}} = 1$  in Section 3.6. We will still focus on results based on  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) in the main text because it is more physically meaningful and contains expected dependencies on  $q$  and  $M_*$ .

When two SMBHs eventually merge, they will also lose mass due to GW radiation. The remnant  $M_{\text{BH}}$  depends on the SMBH properties in a complex manner, and we assume that, typically, 10% of the less-massive  $M_{\text{BH}}$  is lost due to GW radiation (e.g., J. Healy et al. 2014; V. Varma et al. 2019). Such a mass loss is only secondary compared to the merger probability and its large uncertainty. It is also worth noting that the GW radiation has recoil upon merging SMBHs, which might cause some SMBHs to be kicked out of the system (e.g., C. O. Lousto et al. 2012). However, this can hardly affect the overall  $p_{\text{merge}}$  as only  $\lesssim 5\%$  of SMBH binaries in the GW-radiation phase are expected to be ejected (e.g., C. O. Lousto et al. 2012; A. Kulier et al. 2015), but it is possible that the GW-recoiled wandering SMBHs may be dominant in dark-matter halo centers (D. Izquierdo-Villalba et al. 2020).

To summarize, if two galaxies, both of which contain central SMBHs, merge according to the merger trees, we first randomly determine if the two SMBHs will merge with a single Bernoulli realization with a success probability of  $p_{\text{merge}}$ . If a merger is expected, we let it happen immediately and generate a single SMBH with the GW-radiation mass loss subtracted. If a merger does not happen, we keep the more massive SMBH as the central one that will continue to grow in future snapshots and put the other one into the wandering SMBH population that will not grow later, and our assumption that wandering SMBHs do not grow has been attested to be generally appropriate in A. Ricarte et al. (2021b). We acknowledge that there is generally a time delay of several hundred mega-years between galaxy mergers and the final SMBH coalescence (e.g., M. Tremmel et al. 2018b). However, such an effect strongly depends on the physical processes discussed above and is difficult to quantify. It is also challenging to adopt a single value to represent the typical

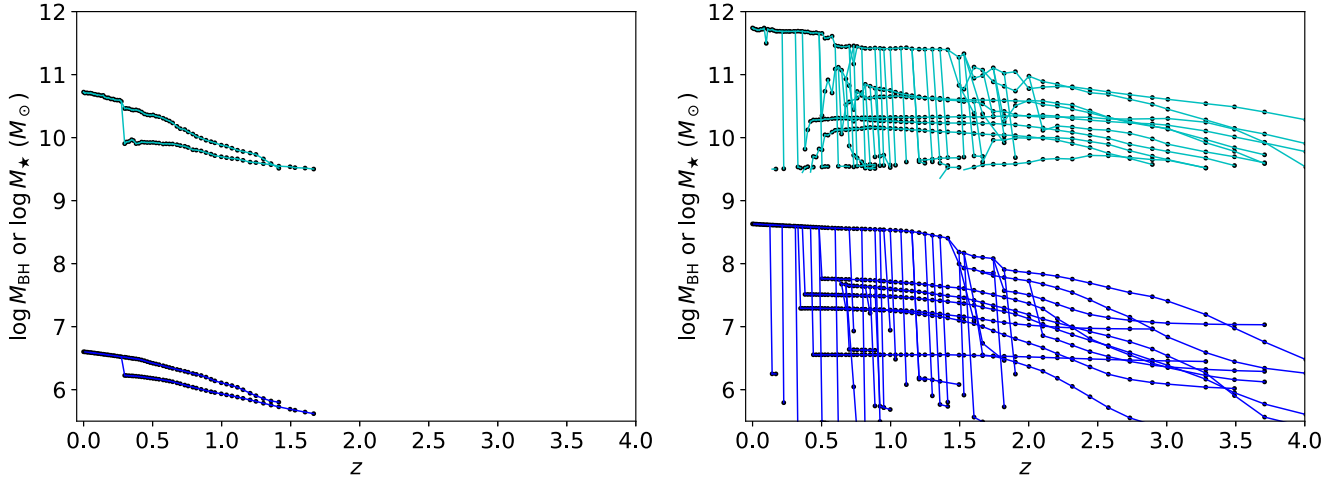
delay time because M. Tremmel et al. (2018b) showed that the delay time is widely distributed, with half of the merged SMBHs having small time delays ( $< 0.5$  Gyr) and the other half distributed in a heavy tail of the delayed time distribution extending to several giga-years. As far as we know, there are actually no well-constructed relations available between the delay time and global galaxy properties. Besides, our TNG snapshots have a time step of 150 Myr, and thus the typical delay timescale corresponds to only a few time steps and is negligible to the overall cosmic time span. Therefore, we do not add any time delays here to the SMBH merger for simplicity. The systematic error caused by this simplification should be smaller than that in the extreme case of  $p_{\text{merge}} = 0$ , which, as we will show in Section 3.6, does not have major effects upon most of our results. Besides, a time delay does not change the overall merged mass but only smears out the mergers over time, and thus it would not cause any difference in population properties if the population mean merger-driven  $M_{\text{BH}}$  growth rate ( $\dot{M}_m$ ) is independent of redshift. As Section 3.3 will show,  $\dot{M}_m$  only weakly correlates with redshift for all the  $M_{\text{BH}}$  or  $M_*$  values, and the changes in  $\dot{M}_m$  within 1 Gyr are  $\lesssim 0.2$  dex; thus delayed mergers will not affect  $\dot{M}_m$  by over 0.2 dex.

## 2.5. TNG

TNG includes ten baryonic simulations in total on three choices of box side lengths, and the flagship, highest-resolution versions of each box side length choice include TNG50-1, TNG100-1, and TNG300-1; see D. Nelson et al. (2019b) for more details. We do not explicitly write the suffix “1” when referring to these flagship TNG simulations hereafter for conciseness. Briefly, these simulations share the same baryonic physics model; side lengths of  $35 h^{-1}$ ,  $75 h^{-1}$ , and  $205 h^{-1}$  (i.e., roughly 50, 100, and 300 for  $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1}) = 0.7$ ) comoving Mpc are used for TNG50, TNG100, and TNG300, respectively, while their baryonic mass resolutions are roughly  $10^5$ ,  $10^6$ , and  $10^7 M_\odot$ , respectively.

As discussed in Section 2.1, we do not directly utilize the SMBH population in TNG. However, we briefly summarize the TNG SMBH modeling for comparison with our own accretion channel. Full details of the TNG SMBHs were described in R. Weinberger et al. (2017, 2018). TNG SMBHs are seeded with  $M_{\text{BH}} = 1.2 \times 10^6 M_\odot$  when dark-matter halos reach  $7.4 \times 10^{10} M_\odot$ . Their accretion rates are assumed to follow the Bondi–Hoyle–Lyttleton model (R. Edgar 2004 and references therein), which is determined by the ambient gas properties, with an upper limit of the Eddington accretion rate. SMBHs further cause feedback in two modes—kinetic and thermal energy is released into the environment when the accretion rates are low and high, respectively. The transition accretion rate between these two modes depends on  $M_{\text{BH}}$  such that more massive SMBHs are more likely to be in the kinetic mode. Regarding the SMBH dynamics, TNG repositions SMBHs in the local potential minimum at each global time step and merges SMBH binaries without a time delay. The simulation parameters are tuned such that the predicted  $M_{\text{BH}}-M_*$  relation at  $z=0$  broadly matches the observed, dynamically measured relation, and several other predicted basic galaxy-population properties can match observations with the SMBH feedback (see R. Weinberger et al. 2017 and A. Pillepich et al. 2018b for more details).

It has been shown that the simulated galaxy population is generally in good agreement with observations (e.g.,



**Figure 2.** Example  $M_*$  and  $M_{\text{BH}}$  evolutionary tracks for MW-like galaxies (left) and massive galaxies with  $M_* \approx 10 M_*^{\text{MW}}$  (right). Each panel presents one galaxy at  $z = 0$  and all of its progenitors with  $M_* > 10^{9.5} M_\odot$  at previous snapshots. Each black point marks one massive galaxy with  $M_* > 10^{9.5} M_\odot$  at the corresponding snapshot, and dwarf galaxies are not shown for visual clarity. Cyan and blue lines connect galaxies with their descendants in terms of  $M_*$  and  $M_{\text{BH}}$ , respectively. A few cyan lines drop down at  $\approx 10^{9.5} M_\odot$  without connecting to descendant points because their descendants have  $M_* < 10^{9.5} M_\odot$  and are thus not shown, but these galaxies will eventually merge to form the final  $z = 0$  galaxy. The massive-galaxy example in the right panel apparently undergoes more mergers with massive galaxies than for the MW-like example in the left panel, which undergoes only one merger with massive galaxies.

A. Pillepich et al. 2018b), such as their SMFs (A. Pillepich et al. 2018a), colors (D. Nelson et al. 2018), and morphology (V. Rodriguez-Gomez et al. 2019). Nevertheless, the consistency, although remarkable, is not perfect. Small differences of  $\lesssim 0.3$  dex exist in terms of the SMF, and the deviations can reach over 1 dex at the very massive end ( $M_* \gtrsim 10^{12} M_\odot$ ), depending upon the  $M_*$  definition in simulations (A. Pillepich et al. 2018a). These values reflect the typical systematic uncertainty of our results originating from simulations. The cosmic variance within finite simulation boxes generally introduce smaller uncertainties. S. Genel et al. (2014) showed that the cosmic variance only introduces an uncertainty of  $\lesssim 0.1$  dex to the SMF at  $z = 0$  for a TNG100 box, and the uncertainty only increases to 0.3 dex in the massive end of the SMF at  $z = 3$ . This cosmic variance will be further suppressed and become negligible for a TNG300 box.

We conduct our analyses for TNG100 and TNG300, both of which have sufficiently large galaxy source statistics. We do not analyze TNG50 for two reasons. First, its volume is small and only contains fewer than 300 galaxies with  $M_* > 10^{10.5} M_\odot$  at  $z = 0$ . Second, it is not necessarily more reliable than TNG100 in terms of the  $M_*$  evolution of massive galaxies. TNG50 was designed for a different purpose of probing the structural and kinematic properties down to a  $\approx 0.1$  kpc scale (D. Nelson et al. 2019a; A. Pillepich et al. 2019), and it is TNG100, not TNG50, that is calibrated against the observed galaxy population. Therefore, TNG50 cannot provide more insights into our specific science than for TNG100.

We adopt the SubLink TNG merger trees derived based on the algorithm in V. Rodriguez-Gomez et al. (2015). We adopt  $M_*$  as the sum of stellar particles within twice the stellar half-mass radius, i.e., `SubhaloMassInRadType[4]` in the TNG catalogs. It is known that simulated galactic properties generally depend on the resolution. For  $M_*$ , it generally increases with better resolution, as explored in, e.g., A. Pillepich et al. (2018a, 2018b). To mitigate this effect, we rescale the TNG300  $M_*$  so that the TNG300 SMF at  $M_* > 10^{9.5} M_\odot$  can match the best with the TNG100 one. We use TNG100 as the reference because it was calibrated with observations, while other TNG simulations are not directly calibrated but share the

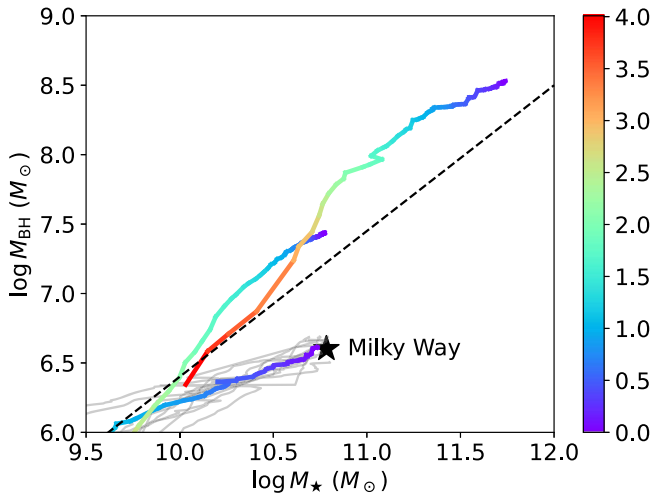
same physics with TNG100. Following the suggestion in A. Pillepich et al. (2018a), we set the rescaling factor to be redshift dependent but not  $M_*$  dependent, which is sufficient for a good correction for the resolution effect, and the typical value is 1.5. Therefore, the resolution rescaling correction is at most 0.2 dex, consistent with the finding in A. Pillepich et al. (2018a).

We select and only analyze galaxies with  $M_* > 10^{9.5} M_\odot$  at  $z = 0$  and their progenitors in the merger trees because, as has been explained in Section 2.2, our understanding of the MBH accretion in dwarf galaxies is still poor. Galaxies not included in merger trees are not considered; such cases are very rare for galaxies with masses far above the mass-resolution limits, and only 1% of galaxies with  $M_* > 10^{9.5} M_\odot$  at  $z = 0$  are excluded. This returns us 10,635 and 186,542 galaxies at  $z = 0$  for TNG100 and TNG300, respectively. The numbers of their progenitors with  $M_* > 10^{9.5} M_\odot$  at  $z = 1, 2, 3$ , and 4 are 8615 (156,018), 5393 (101,663), 2553 (49,630), and 970 (19,355) for TNG100 (TNG300), respectively. To keep the narrative flow concise, we focus on TNG300 in the main text because it has better number statistics than TNG100. We will not explicitly write “TNG300” hereafter, and unless noted in the main text, we always refer to TNG300. TNG100 and TNG300 results are generally similar, as illustrated in Appendix B.

### 3. Results

Before analyzing the overall population, we show typical example evolution results for galaxies similar to the Milky Way (MW) and more massive ones to provide first insights. The MW has  $M_*^{\text{MW}} = 6.08 \times 10^{10} M_\odot$  (e.g., T. C. Licquia & J. A. Newman 2015) and  $M_{\text{BH}}^{\text{MW}} = 4.0 \times 10^6 M_\odot$  (e.g., Event Horizon Telescope Collaboration et al. 2022), and we define MW-like galaxies as those within  $\log M_*^{\text{MW}} \pm 0.1$  dex and  $\log M_{\text{BH}}^{\text{MW}} \pm 0.1$  dex at  $z = 0$ . We also select massive galaxies with  $M_*$  higher than  $M_*^{\text{MW}}$  by 0.9–1.1 dex. Figure 2 presents examples of the  $M_*$  and  $M_{\text{BH}}$  evolution for MW-like galaxies and massive ones. MW-like galaxies typically reach our  $M_*$  threshold of  $10^{9.5} M_\odot$  at intermediate redshifts of  $\approx 1.5$ –2 and undergo few mergers with massive galaxies, while the massive





**Figure 3.** Example evolutionary tracks on the  $M_{\text{BH}}-M_*$  plane. The black star marks the MW. The gray curves represent individual MW-like galaxies. The other three multicolor curves are color-coded by the redshift. The one connected to the black star is the median track of MW-like galaxies, and the other two are the median tracks of galaxies with  $M_* \approx M_*^{\text{MW}}$  and  $M_* \approx 10M_*^{\text{MW}}$ . The dashed line is the AGN-based  $M_{\text{BH}}-M_*$  relation in A. E. Reines & M. Volonteri (2015).

ones are already above the  $M_*$  threshold at  $z=4$  and undergo many mergers during their evolution. Figure 3 further shows their evolutionary tracks in the  $M_{\text{BH}}-M_*$  plane. The MW hosts a smaller SMBH compared to the median  $M_{\text{BH}}$  of other galaxies with similar  $M_*$  by a factor of 7 (0.8 dex), and thus the tracks of our MW-like galaxies are generally below the  $M_{\text{BH}}-M_*$  relation in A. E. Reines & M. Volonteri (2015). Compared to other galaxies with similar  $M_*$ , the  $M_*$  growth of these MW-like galaxies happens at lower redshifts. This is required to produce a smaller  $M_{\text{BH}}$  in our methodology. Since  $\dot{M}_a$  increases with rising redshift at a given  $M_*$ , the later  $M_*$  growth of MW-like galaxies can lead to smaller  $\dot{M}_a$ . We note that the MW does not stand out as a strong outlier because the  $M_{\text{BH}}$  scatter of our sources with  $M_* \approx M_*^{\text{MW}}$  is 0.4 dex due to the diverse growth history of  $M_*$ , and this scatter is only a lower limit because the  $(M_*, z)$  predictor set does not capture all the variance of  $\dot{M}_a$  (Section 3.1). We do not attempt to link the mass assemblies of these MW-like galaxies to the mass assembly of the actual MW, which has been an open question for decades. For example, we compared the estimated MW mass-assembly histories in different literature works (e.g., N. J. Fantin et al. 2019; M. Xiang & H.-W. Rix 2022) with those of MW-like galaxies in P. G. van Dokkum et al. (2013) and did not find consistent evidence suggesting the MW mass assembly is relatively earlier or later. The MW SMBH may also have undergone a previous major merger (e.g., Y. Wang & B. Zhang 2024) that further complicates the growth history. Instead, we solely use the examples above to illustrate how a typical SMBH evolves together with its host galaxy and how different  $M_*$  evolutionary tracks can lead to different  $M_{\text{BH}}$ .

We will now present the overall population analyses in the following subsections. Specifically, Section 3.1 presents the  $M_{\text{BH}}-M_*$  scaling relation and its cosmic evolution; Section 3.2 presents the BHMF and its cosmic evolution; Section 3.3 analyzes the growth rates in the accretion and merger channels; Section 3.4 presents the overall redshift evolution of  $M_{\text{BH}}$ ; Section 3.5 discusses wandering SMBHs; Section 3.6 discusses the impacts of the  $p_{\text{merge}}$  choice on our results; and Section 3.7

discusses the impacts of the SMBH seeding choice on our results.

### 3.1. The $M_{\text{BH}}-M_*$ Relation

$M_{\text{BH}}$  has been shown to follow a correlation with the host bulge mass and also  $M_*$ , which is an indication of the coevolution between SMBHs and galaxies (e.g., J. Kormendy & L. C. Ho 2013 and references therein). Although the correlation with the bulge is tighter, we focus on  $M_*$  because our whole methodology is integrated based on  $M_*$ .

Table 1 presents our predicted  $M_{\text{BH}}-M_*$  relations at different redshifts based on both TNG100 and TNG300. We plot the TNG300 results in the top panel of Figure 4 and also show the original TNG-simulated  $z=0$  relation and the relations in J. Kormendy & L. C. Ho (2013), A. E. Reines & M. Volonteri (2015), F. Shankar et al. (2016), G. A. D. Savorgnan et al. (2016), J. E. Greene et al. (2020), and J. I. H. Li et al. (2023) for comparison, where the early-type relation is plotted for G. A. D. Savorgnan et al. (2016), and the relation combining both early- and late-type galaxies is adopted for J. E. Greene et al. (2020). These literature works were based on different kinds of  $M_{\text{BH}}$  measurements. J. Kormendy & L. C. Ho (2013), G. A. D. Savorgnan et al. (2016), J. E. Greene et al. (2020), and the higher relation in A. E. Reines & M. Volonteri (2015) are derived from dynamical measurements, while J. I. H. Li et al. (2023) and the lower relation in A. E. Reines & M. Volonteri (2015) are from broad-line AGNs. It is apparent that the dynamically measured ones are generally higher than the AGN-based ones. F. Shankar et al. (2016) argued that the main reason is that the SMBH sample with dynamical  $M_{\text{BH}}$  measurements is biased toward more massive SMBHs, and the higher scaling relations are convolution results of the *debiased* scaling relation, which is also plotted in Figure 4 and closer to the AGN-based relations, with selection biases. It is also worth noting that the highest three relations (J. Kormendy & L. C. Ho 2013; A. E. Reines & M. Volonteri 2015; G. A. D. Savorgnan et al. 2016) in Figure 4 are all from early-type galaxies. Late-type galaxies tend to have lower  $M_{\text{BH}}$  at a given  $M_*$  (e.g., G. A. D. Savorgnan et al. 2016; J. E. Greene et al. 2020). We plot the dynamically measured scaling relation combining both early- and late-type galaxies from J. E. Greene et al. (2020) in Figure 4, and its normalization is  $>0.5$  dex lower than for early-type galaxies, which helps mitigate the difference between dynamically measured scaling relations and AGN-based ones.

The figure shows that our  $z=0$  scaling relation is more similar to the AGN-based ones. This is not due to the fact that our initial SMBH seeding follows the AGN-based relation in A. E. Reines & M. Volonteri (2015). We have tried adopting the higher dynamically measured relation as the seeding relation, and the recovered scaling relation is still below the dynamically measured one, as will be shown in Section 3.7. The main reason is that, even if we begin with the dynamically measured relation, the subsequent accretion is not sufficient to maintain the high normalization. Therefore, our X-ray-based  $\dot{M}_a$  from F. Zou et al. (2024) is inconsistent with the dynamically measured relation. The same finding was also shown in F. Shankar et al. (2020b). This may be because X-ray surveys miss some accretion power or the scaling relation is overestimated. However, to match the scaling relation by increasing  $\dot{M}_a$ , we need to elevate it by 1 order of magnitude. We do not think this level of uncertainty can be from

**Table 1**  
The  $M_{\text{BH}}-M_*$  Scaling Relation

$\log M_*$	$z = 0$	$z = 0.5$	$z = 1$	$z = 1.5$	$z = 2$	$z = 2.5$	$z = 3$	$z = 3.5$	$z = 4$
9.50–9.75	5.99	6.00	6.02	6.00	5.96	5.96	5.99	5.95	6.00
	5.99	5.99	6.02	5.99	5.96	5.95	5.96	5.96	5.98
9.75–10.00	6.17	6.28	6.34	6.29	6.17	6.13	6.08	6.12	6.24
	6.17	6.26	6.33	6.29	6.16	6.11	6.09	6.11	6.25
10.00–10.25	6.44	6.60	6.71	6.76	6.62	6.46	6.41	6.44	6.54
	6.43	6.55	6.69	6.79	6.58	6.43	6.38	6.42	6.52
10.25–10.50	6.84	6.96	7.14	7.21	7.06	6.90	6.84	6.89	6.80
	6.75	6.88	7.10	7.18	7.04	6.85	6.82	6.84	6.76
10.50–10.75	7.22	7.34	7.46	7.54	7.48	7.35	7.35	7.26	7.14
	7.21	7.32	7.45	7.55	7.46	7.34	7.37	7.27	7.01
10.75–11.00	7.52	7.63	7.75	7.80	7.79	7.75	7.78	7.62	...
	7.56	7.65	7.79	7.82	7.78	7.71	7.74	7.57	7.27
11.00–11.25	7.86	7.96	8.07	8.05	8.19	8.12	8.04	...	...
	7.91	7.99	8.10	8.16	8.11	8.09	8.05	7.88	7.52
11.25–11.50	8.10	8.23	8.29	8.41	...	...	...	...	...
	8.18	8.29	8.43	8.54	8.54	8.43	8.36	8.15	7.81
11.50–11.75	8.36	8.55	...	...	...	...	...	...	...
	8.49	8.59	8.79	8.93	8.94	8.83	8.65	...	...
11.75–12.00	...	...	...	...	...	...	...	...	...
	8.82	9.00	9.16	9.37	...	...	...	...	...

**Note.** The first column lists the  $\log M_*$  range of each bin, and the other columns show the median  $\log M_{\text{BH}}$  values at the corresponding bins. Values in bins with fewer than 20 sources are not shown. Within each cell, the top and bottom parts show the TNG100 and TNG300 results, respectively.

bolometric corrections or radiative efficiency (Section 2.3), and it is hard to explain how the majority of the accretion power could be missed by X-rays. That said, a more plausible reason, as proposed in F. Shankar et al. (2016, 2020b, 2020a), is that the dynamically measured SMBH sample is biased toward more massive SMBHs. Nevertheless, it should be noted that this claim is still under debate, and some works also argued that the AGN-based relation in A. E. Reines & M. Volonteri (2015) or even the higher dynamically measured one may underestimate the number of SMBHs at the massive end (e.g., D. W. Pesce et al. 2021; G. Sato-Polito et al. 2023).

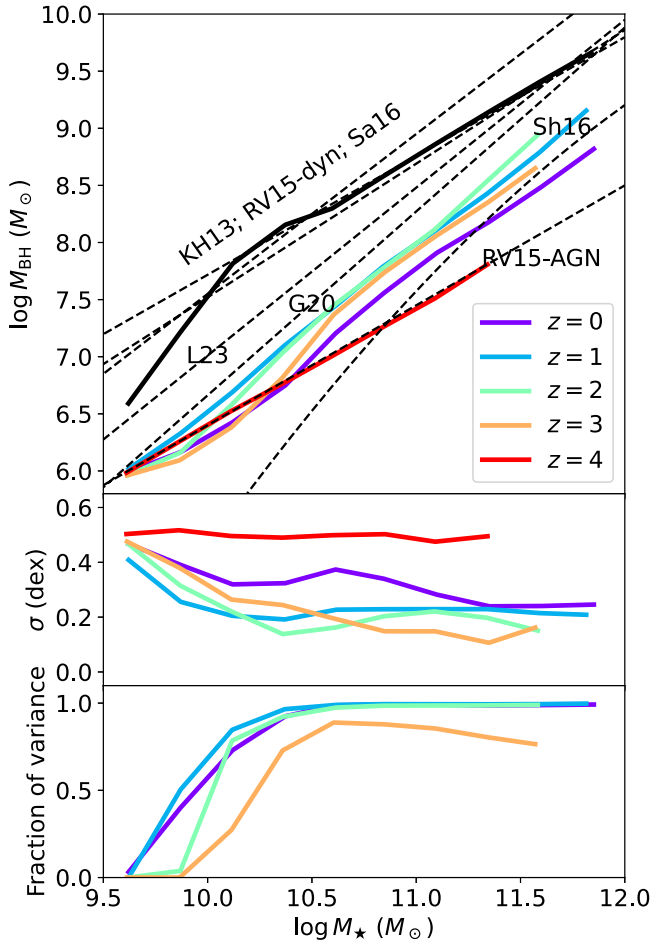
Besides, the original TNG-simulated relation is consistent with the higher dynamically measured relations, which is expected because TNG is tuned to match these relations. The discrepancy between our post-processed results and the original TNG ones originates from the fact that the simulated  $\dot{M}_a$  is greatly different from observations. We compare the original simulated  $\dot{M}_a$ , derived by averaging the `SubhaloBHMDot` column in the TNG catalogs in each  $M_*$  bin, and the observed  $\dot{M}_a$  in F. Zou et al. (2024) as a function of  $M_*$  at different redshifts in Figure 5. The figure reveals that TNG can overestimate  $\dot{M}_a$  by orders of magnitude, especially at  $M_* \lesssim 10^{11} M_\odot$ . Besides, the simulated  $\dot{M}_a$  does not further increase with  $M_*$  at  $M_* \gtrsim 10^{10.5} M_\odot$  and  $z \lesssim 2$  due to the suppression of accretion (R. Weinberger et al. 2018), which contradicts observations. Therefore, even if the original TNG simulations can recover a plausible  $M_{\text{BH}}-M_*$  scaling relation, which is an integrated metric, its simulated accretion channel is still generally unrealistic (M. Habouzit et al. 2021, 2022). This is not surprising because TNG was not calibrated with the AGN population. Given this, our post-processing of the TNG data with direct observational constraints added can improve the SMBH population compared to the original purely simulation-based results. Also, the comparison here is mainly qualitative, and more quantitative comparisons require constructing detailed mock observations from the TNG simulations to convert the TNG results into the space of direct observables (e.g.,

M. Habouzit et al. 2022), which is beyond the scope of this work. Besides, note that the limitation of the simulated accretion does not directly propagate into the simulated galaxy properties in the sense that galaxies are instead regulated by the feedback, whose parameters can be tuned independently from the accretion model. Given the fact that the simulated galaxy population has been calibrated with several metrics (e.g., the SMF, the cosmic star formation rate density, the correlation between  $M_*$  and halo mass, the halo gas fraction, and the galaxy size; see A. Pillepich et al. 2018b), the simulated galaxies are generally more reliable than the simulated SMBHs.

Our scaling relation in Figure 4 does not evolve much from  $z = 4$  to  $z = 0$ , with a difference  $\lesssim 0.5$  dex. The  $z = 0$  curve is slightly systematically below the higher-redshift ones, but the difference is still small ( $< 0.3$  dex). Such small evolution is generally supported by recent works, including both observational studies and cosmological simulations (e.g., M. Sun et al. 2015; G. Yang et al. 2018b, 2019; F. Shankar et al. 2020b; H. Suh et al. 2020; J. Li et al. 2021a; J. I. H. Li et al. 2021b; M. Habouzit et al. 2021; G. Yang et al. 2022; J. I. H. Li et al. 2023).

Some parts of the scaling relations are dominated by our initial SMBH seeding; for example, our  $z = 4$  curve, by design, exactly follows A. E. Reines & M. Volonteri (2015) with perturbations from the seeding scatter. The contributions of our adopted seed mass to  $M_{\text{BH}}$  will be explored in greater detail in Section 3.4. Briefly, SMBHs in massive galaxies with  $M_* \gtrsim 10^{10} M_\odot$  at  $z = 3$  have already lost their memory about their seed masses at  $z = 4$ , while SMBHs in low-mass galaxies with  $M_* \lesssim 10^{10} M_\odot$  at  $z = 0$  still retain significant information at high redshifts. Therefore, the low-mass parts of our scaling relations in Figure 4 are mainly set by our initial conditions at all redshifts, while the more massive parts ( $M_* \gtrsim 10^{10} M_\odot$ ) at  $z \leq 3$  are instead genuinely created by the SMBH growth. Therefore, our results in the massive parts should be robust as long as the accumulated SMBH masses before  $z = 4$  are negligible. We will present the SMBH growth rates via different channels in Section 3.3, where we will show that the SMBH growth is

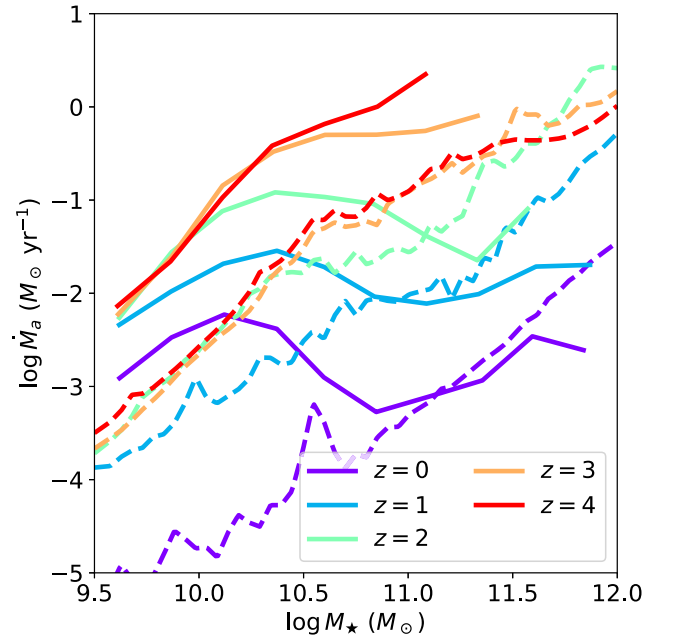




**Figure 4.** Top: our  $M_{\text{BH}}-M_*$  relations at different redshifts, plotted as the colored curves. The original TNG-simulated  $z=0$  relation is shown as the black solid curve for comparison. We also plot the relations in J. Kormendy & L. C. Ho (2013), J. I. H. Li et al. (2023), J. E. Greene et al. (2020), A. E. Reines & M. Volonteri (2015), F. Shankar et al. (2016), and G. A. D. Savorgnan et al. (2016) as the black dashed lines, which are labeled as KH13, L23, G20, RV15, Sh16, and Sa16, respectively. Two relations in A. E. Reines & M. Volonteri (2015) are included—one based on the SMBH sample with dynamically measured  $M_{\text{BH}}$ , and the other based on the AGN sample; these two are labeled as RV15-dyn and RV15-AGN, respectively. The highest three relations are labeled together as “KH13; RV15-dyn; Sa16” for a better visualization, and they are all from dynamically measured SMBHs in early-type galaxies. Our recovered relations are apparently lower than the original TNG relation and the early-type dynamically measured ones. Middle: our recovered  $1\sigma$   $M_{\text{BH}}$  scatter in dex vs.  $M_*$ . Bottom: the fractional contribution of SMBH growth along different  $M_*$  evolutionary tracks to our recovered variance, which is one minus the fraction of variance from seeding scatter.

mainly dominated by the accretion channel in most cases, and thus our scaling relations above  $M_* \gtrsim 10^{10} M_\odot$  are driven by the correlation between  $\dot{M}_a$  and  $(M_*, z)$ . This is consistent with G. Yang et al. (2018b) and F. Shankar et al. (2020b), who neglected the merger channel, integrated  $\dot{M}_a(M_*, z)$ , and found that the local-Universe scaling relation can apparently be recovered and is close to the lower, AGN-based scaling relation.

It is also worth noting that, if we adopt a higher seeding relation such as the dynamically measured ones, the recovered scaling relation would increase with increasing redshift, as will be shown in more detail in Section 3.7. This is because the high seeding scaling-relation normalization cannot be maintained by the subsequent accretion, and thus the overall normalization gradually decreases as the Universe ages. However, such a redshift-dependent evolution requires SMBHs to accumulate at least half of their masses by  $z=4$  (see Section 3.7), which



**Figure 5.**  $\dot{M}_a$  vs.  $M_*$  at different redshifts. The solid and dashed curves are from the original TNG300 simulation and the observed relation in F. Zou et al. (2024), respectively. There are orders-of-magnitude differences between the simulated and observed  $\dot{M}_a$ , and the simulated  $\dot{M}_a$  often lacks a positive correlation with  $M_*$ .

contradicts previous findings that the accumulation of half of the  $M_{\text{BH}}$  occurred much later; see F. Vito et al. (2016, 2018), F. Shankar et al. (2020b), X. Shen et al. (2020), M. Habouzit et al. (2021), and H. Zhang et al. (2023) for examples from a variety of observational or theoretical perspectives; but also note that recent JWST results tend to favor stronger early SMBH growth with over-massive SMBHs at  $z > 4$  (e.g., F. Pacucci et al. 2023; R. Maiolino et al. 2023). This redshift dependence of the scaling relation also contradicts observations that no apparent evolution is seen with respect to redshift (e.g., Figure 10 in J. I. H. Li et al. 2023).

In the middle panel of Figure 4, we plot our recovered  $1\sigma$  scatter of the scaling relation versus  $M_*$ . Our recovered  $M_{\text{BH}}$  scatter is contributed by two components—one is from the seeding scatter, and the other is from the SMBH growth along diverse  $M_*$  evolutionary tracks. To assess the relative contributions from these two components, we rerun our analyzes in Section 2 but set  $\dot{M}_a = 0$ , and the resulting  $M_{\text{BH}}$  only keeps the seeding information, as will be elaborated more in Section 3.4. We run it twice so that we can calculate the contribution of the seeding scatter to the total  $M_{\text{BH}}$  variance based on the dispersion of the difference in  $M_{\text{BH}}(\dot{M}_a = 0)$  between these two independent runs.<sup>14</sup> We then present the fractional contribution of variance from the second component,

<sup>14</sup> Mathematically,  $M_{\text{BH}}(\dot{M}_a = 0) = \overline{M_{\text{BH}}}(\dot{M}_a = 0) + \varepsilon_{\text{seed}}$ , where the first term is the mean  $M_{\text{BH}}(\dot{M}_a = 0)$ , and the second term is a random variable with a mean of 0 originating from the seeding scatter. We do not directly assess  $\overline{M_{\text{BH}}}(\dot{M}_a = 0)$  because it cannot be measured with a simple run without seeding scatter. The reason is that the seeding scatter is added in the logarithmic space, and eliminating the scatter would change  $\overline{M_{\text{BH}}}(\dot{M}_a = 0)$  due to the fact that the mean of a log-normal distribution centered at 0 is not 0. Instead, we use two independent runs with seeding scatters to subtract this term. For these two realizations, we have  $\delta M_{\text{BH}} = M_{\text{BH}}(\text{realization 1}) - M_{\text{BH}}(\text{realization 2}) = \varepsilon_{\text{seed}}(\text{realization 1}) - \varepsilon_{\text{seed}}(\text{realization 2})$ . Thus,  $\text{Var}(\delta M_{\text{BH}}) = 2\text{Var}(\varepsilon_{\text{seed}})$ . Each SMBH has a  $\delta M_{\text{BH}}$  value that can be regarded as one realization of the underlying population  $\delta M_{\text{BH}}$  distribution. Thus, the typical contribution from the seeding scatter to  $M_{\text{BH}}$ ,  $\text{Var}(\varepsilon_{\text{seed}})$ , can be measured from the dispersion of  $\delta M_{\text{BH}}$ .

**Table 2**  
The BHMF

$\log M_{\text{BH}}$	$z = 0$	$z = 0.5$	$z = 1$	$z = 1.5$	$z = 2$	$z = 2.5$	$z = 3$	$z = 3.5$	$z = 4$
6.00–6.25	–2.28 –2.39	–2.33 –2.42	–2.41 –2.50	–2.50 –2.58	–2.58 –2.62	–2.68 –2.72	–2.81 –2.86	–3.01 –3.07	–3.30 –3.31
6.25–6.50	–2.26 –2.34	–2.30 –2.36	–2.39 –2.44	–2.45 –2.54	–2.57 –2.63	–2.65 –2.70	–2.85 –2.88	–3.10 –3.09	–3.31 –3.32
6.50–6.75	–2.33 –2.38	–2.30 –2.37	–2.39 –2.45	–2.54 –2.61	–2.65 –2.69	–2.76 –2.80	–2.98 –2.99	–3.17 –3.21	–3.40 –3.37
6.75–7.00	–2.44 –2.50	–2.45 –2.51	–2.49 –2.54	–2.58 –2.60	–2.72 –2.75	–2.93 –2.96	–3.11 –3.17	–3.35 –3.34	–3.43 –3.51
7.00–7.25	–2.59 –2.63	–2.57 –2.60	–2.56 –2.57	–2.67 –2.69	–2.84 –2.87	–3.10 –3.12	–3.33 –3.36	–3.48 –3.52	–3.71 –3.70
7.25–7.50	–2.69 –2.71	–2.65 –2.68	–2.62 –2.68	–2.71 –2.75	–2.94 –2.96	–3.24 –3.25	–3.56 –3.47	–3.80 –3.64	–4.07 –3.91
7.50–7.75	–2.84 –2.84	–2.79 –2.80	–2.81 –2.83	–2.84 –2.88	–3.08 –3.09	–3.46 –3.37	–3.73 –3.61	–3.95 –3.86	... –4.19
7.75–8.00	–3.03 –2.98	–2.96 –2.93	–2.99 –2.94	–3.16 –3.06	–3.45 –3.29	–3.78 –3.60	–3.92 –3.79	... –4.20	... –4.52
8.00–8.25	–3.33 –3.25	–3.33 –3.22	–3.39 –3.27	–3.57 –3.42	–3.78 –3.64	–3.97 –3.86	... –4.17	... –4.49	... –4.94
8.25–8.50	–3.69 –3.62	–3.73 –3.61	–3.76 –3.66	–3.85 –3.79	–4.09 –4.04	... –4.24	... –4.40	... –4.94	... –5.23
8.50–8.75	–4.07 –3.96	–4.07 –4.00	–4.15 –4.08	... –4.25	... –4.38	... –4.66	... –4.88	... ...	... ...
8.75–9.00	... –4.33	... –4.34	... –4.40	... –4.56	... –4.77	... –4.98	... ...	... ...	... ...
9.00–9.25	... –4.60	... –4.67	... –4.79	... –4.97	... –5.06	... ...	... ...	... ...	... ...
9.25–9.50	... –4.97	... –5.02	... –5.19	... –5.40	... ...	... ...	... ...	... ...	... ...
9.50–9.75	... –5.23	... –5.32	... –5.42	... ...	... ...	... ...	... ...	... ...	... ...

**Note.** The first column lists the  $\log M_{\text{BH}}$  range of each bin, and the other columns show  $\log \phi(M_{\text{BH}})$  in  $\text{Mpc}^{-3} \text{dex}^{-1}$  in the corresponding bins. Values in bins with fewer than 20 sources are not shown. Within each cell, the top and bottom parts show the TNG100 and TNG300 results, respectively.

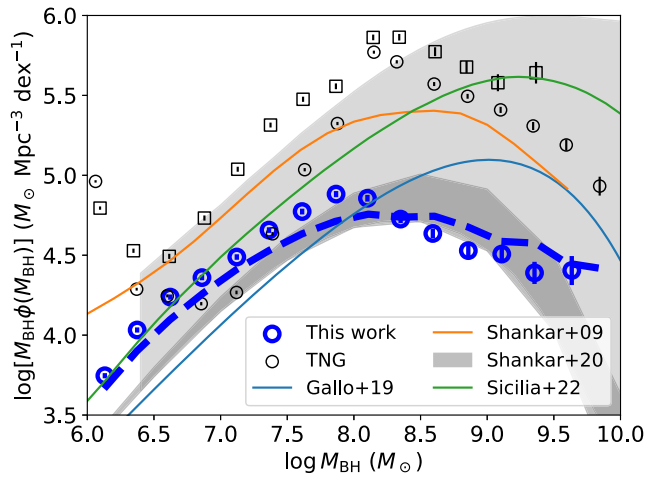
the SMBH growth along diverse  $M_*$  evolutionary tracks, in the bottom panel of Figure 4. The figure shows that our recovered variance is dominated by this second component at  $M_* > 10^{10} M_\odot$  and  $z < 3$ , indicating the memory retained about  $M_{\text{BH}}$  seed masses quickly fades away as  $M_*$  increases from  $10^{9.5} M_\odot$  to  $10^{10} M_\odot$ . We focus on the low-memory regime here, where our predicted local-Universe scatter is  $\approx 0.3 - 0.4$  dex. The observationally measured intrinsic scatter is  $\approx 0.5$  dex (A. E. Reines & M. Volonteri 2015; J. I. H. Li et al. 2023). It should be noted that our scatters are only lower limits because we assumed that all the SMBHs exactly follow the  $\dot{M}_a(M_*, z)$  function; however, there is an additional variance in  $\dot{M}_a$  not captured by the  $(M_*, z)$  predictor set, such as the residual dependence on star formation rate and morphology (e.g., compactness). Nevertheless, the ratio between our variance and the observed one reaches  $\approx 50\%$ , indicating that the SMBH growth along diverse  $M_*$  evolutionary tracks can already explain half of the observed  $M_{\text{BH}} - M_*$  variance. The middle panel of Figure 4 also indicates that the scatter may decrease from  $z = 0$  to  $z = 1$  and remains roughly constant up to  $z = 3$ , but this conclusion is subject to the amount of variance not covered by our  $\dot{M}_a(M_*, z)$  function.

### 3.2. BHMF

The BHMF, denoted as  $\phi(M_{\text{BH}})$ , is the number of SMBHs per unit comoving volume per unit  $\log M_{\text{BH}}$ . Although this quantity is fundamental, it is a nontrivial challenge to measure

it reliably because indirect tracers of  $M_{\text{BH}}$  are needed; see, for example, B. C. Kelly & A. Merloni (2012) for a review. With our constructed SMBH population, we can compute the BHMF at different redshifts. We do not include wandering SMBHs in our BHMF because usually only central SMBHs are observable, and the BHMF in previous literature also refers to central SMBHs. We will focus on these wandering SMBHs in Section 3.5. Given our  $M_*$  cut at  $10^{9.5} M_\odot$ , which corresponds to  $M_{\text{BH}} = 7 \times 10^5 M_\odot$  according to the scaling relation in A. E. Reines & M. Volonteri (2015), we focus on the BHMF with  $M_{\text{BH}} > 10^6 M_\odot$ . It is possible that some dwarf galaxies contain SMBHs above this threshold, and thus our BHMF may be slightly underestimated at low  $M_{\text{BH}}$ , but this bias should be negligible when  $M_{\text{BH}} \gtrsim 10^7 M_\odot$ .

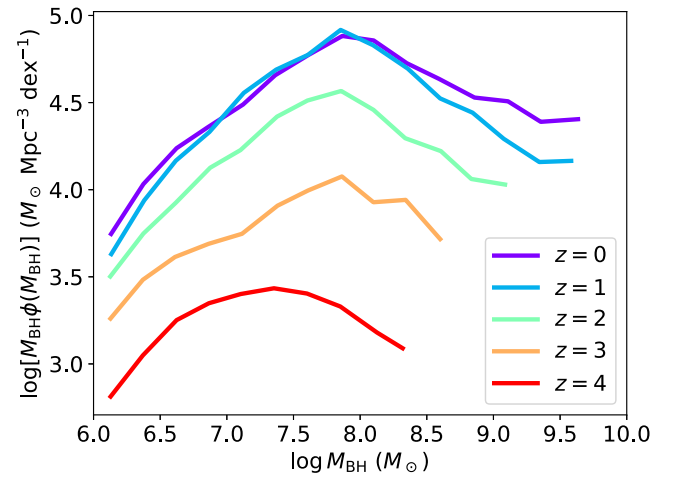
We bin our sources with an  $M_{\text{BH}}$  step of 0.25 dex to measure the BHMF, and only those bins with at least 20 sources are used. The BHMF at different redshifts is tabulated in Table 2 for both TNG100 and TNG300. Figure 6 presents  $M_{\text{BH}}\phi(M_{\text{BH}})$  at  $z = 0$ . We show  $M_{\text{BH}}\phi(M_{\text{BH}})$ , the cosmic SMBH mass density per unit  $\log M_{\text{BH}}$ , instead of  $\phi(M_{\text{BH}})$  to illustrate the typical  $M_{\text{BH}}$  range where most SMBH mass resides. The massive end of the BHMF may visually look like a plateau, which is primarily due to statistical fluctuations. If we further plot the BHMF to higher masses, the BHMF would continue to decrease. However, it is worth noting that our massive end generally does not decay as fast as that of other works. This originates from the fact that the TNG SMF also does not decay sufficiently fast at the very massive end (A. Pillepich et al. 2018a). Our uncertainties are



**Figure 6.** The BHMf at  $z = 0$ . We show  $M_{\text{BH}}\phi(M_{\text{BH}})$  instead of  $\phi(M_{\text{BH}})$  to illustrate the typical  $M_{\text{BH}}$  range where most SMBH mass resides. The blue open circles with  $1\sigma$  error bars are our BHMf, where the uncertainties only represent the limited number of sources included in the simulation boxes. The blue dashed line is the resulting BHMf with an additional scatter of 0.5 dex added to our  $M_{\text{BH}}$ . The black open circles and squares with  $1\sigma$  error bars are the BHMf in the original TNG300 and TNG100 simulations. The BHMf in E. Gallo & A. Sesana (2019), A. Sicilia et al. (2022), and F. Shankar et al. (2020a, 2009) are also plotted, as labeled in the legend. The lighter shaded region shows the uncertainty range in F. Shankar et al. (2020a) considering the uncertainties of scaling relations between  $M_{\text{BH}}$  and galaxy properties (see also A. Sicilia et al. 2022), and the darker shaded region is the BHMf in F. Shankar et al. (2020a) based on the scaling relation after correcting for the selection bias of dynamically measured SMBHs.

calculated as the Poisson error of the number of sources in each bin following N. Gehrels (1986) and do not include other systematic uncertainties, which could be larger. There are also additional uncertainties not accounted for from the statistical uncertainties of the  $\dot{M}_a(M_*, z)$  function, but Appendix A shows that these terms should be small in general. There are several sources of systematic uncertainties. For example, as mentioned in Section 2.3, there is a typical systematic uncertainty of up to a factor of a few when measuring  $\dot{M}_a$  from X-ray emission. Another source of systematic uncertainties is from TNG, i.e., how real the simulated universe is. Different models can produce different evolutionary tracks of  $M_*$  for galaxies (e.g., see Figure 3 in F. Shankar et al. 2020b). Besides,  $\dot{M}_a(M_*, z)$  only captures part of the variance in SMBH growth, and our recovered scatter of the  $M_{\text{BH}}-M_*$  scaling relation in Section 3.1 is smaller than the observed one. We also tried adding an additional scatter of 0.5 dex to  $M_{\text{BH}}$  and systematically reducing all  $M_{\text{BH}}$  by 0.29 dex to keep the total mass unchanged.<sup>15</sup> This additional scatter increases the total scatter of  $M_{\text{BH}}$  at a given  $M_*$  to be 0.6–0.7 dex. The new BHMf is plotted as the blue dashed curve in Figure 6, which presents a slightly heavier massive tail.

We also show the BHMf in F. Shankar et al. (2009), E. Gallo & A. Sesana (2019), F. Shankar et al. (2020a), A. Sicilia et al. (2022), and the original TNG100 and TNG300 simulations as comparisons. The former three are all derived from scaling relations of  $M_{\text{BH}}$  with galaxy properties, while the latter three are instead theoretical ones. The lighter shaded region in Figure 6 is from Shankar et al. (2020a, see also A. Sicilia et al. 2022) and reflects the overall uncertainty of the scaling relation and also sub-dominant Poisson uncertainties of



**Figure 7.**  $M_{\text{BH}}\phi(M_{\text{BH}})$  at different redshifts.

limited source counts. Its upper bound is based on the dynamically measured scaling relation, while its lower bound is derived by correcting the selection bias that more massive SMBHs are more likely to have dynamically measured  $M_{\text{BH}}$ , which, hence, contains an assumption that the observed scaling relation is highly biased, as discussed in Section 3.1. We also explicitly plot the debiased BHMf in F. Shankar et al. (2020a) as the darker shaded region in Figure 6. The typical BHMf uncertainty range originating from the scaling relation (i.e., biased or not) is  $\approx 1$  dex below  $10^8 M_\odot$  but increases to several orders of magnitude at the massive end where the BHMf exponentially decays.

Our BHMf is more consistent with the debiased BHMf in F. Shankar et al. (2020a), echoing our finding in Section 3.1 that our recovered scaling relation is lower than the dynamically measured relation. However, as Section 3.7 will show, the BHMf could be elevated, especially at the massive end, if we seed SMBHs with higher  $M_{\text{BH}}$ , but this requires the SMBH growth before  $z = 4$  to have a considerable fractional contribution to the final  $M_{\text{BH}}$  at  $z = 0$ . Our BHMf also shows a notable difference from the original TNG-simulated SMBHs, as similarly shown in Section 3.1.

Integrating the BHMf, we obtain the local SMBH mass density to be  $\rho_{\text{BH}} = 1.3 \times 10^5 M_\odot \text{Mpc}^{-3}$  and  $1.5 \times 10^5 M_\odot \text{Mpc}^{-3}$  for TNG100 and TNG300, respectively. These numbers are in agreement with recent findings in F. Shankar et al. (2020a), who derived  $\rho_{\text{BH}} = (1-2) \times 10^5 M_\odot \text{Mpc}^{-3}$ . Earlier results based on the local scaling relations between  $M_{\text{BH}}$  and galaxy properties, which tend to be higher than the debiased relation, may return slightly larger  $\rho_{\text{BH}} = (3-6) \times 10^5 M_\odot \text{Mpc}^{-3}$  (e.g., A. W. Graham & S. P. Driver 2007; W. N. Brandt & D. M. Alexander 2015 and references therein). By assuming a radiative efficiency of 0.1 and integrating the bolometric quasar luminosity function from high redshifts to  $z = 0$ , P. F. Hopkins et al. (2007) and X. Shen et al. (2020) derived  $\rho_{\text{BH}} = 5 \times 10^5 M_\odot \text{Mpc}^{-3}$ . We also integrated the AGN X-ray luminosity function in Y. Ueda et al. (2014), adopting the bolometric correction from the X-ray luminosity to the bolometric luminosity in F. Duras et al. (2020), and obtained a lower  $\rho_{\text{BH}} = (2-3) \times 10^5 M_\odot \text{Mpc}^{-3}$ . Therefore, there is an uncertainty of a factor of a few in  $\rho_{\text{BH}}$ , depending upon the detailed methodology. Our values, similar to F. Shankar et al. (2020a), are close to the lower bound compared to previous works.

We further explore the redshift evolution of the BHMf in Figure 7. The BHMf at  $z = 4$  reflects our initial condition and

<sup>15</sup> To ensure the mean of a log-normal distribution is 0, its center should be negative, as reflected by our leftward shift of  $M_{\text{BH}}$ .



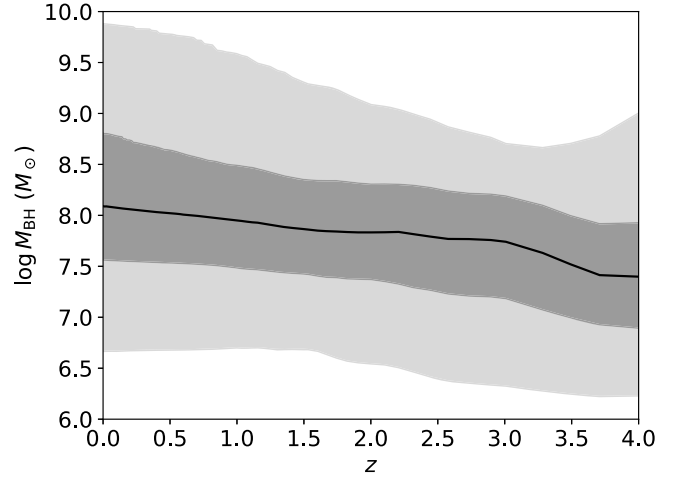
**Table 3**  
Typical  $M_{\text{BH}}$  at Different Redshifts

$z$	5%	25%	50%	75%	95%
0.0	6.52	7.37	7.88	8.44	9.28
	6.67	7.56	8.09	8.81	9.88
0.5	6.55	7.36	7.83	8.29	9.11
	6.68	7.53	8.02	8.64	9.78
1.0	6.57	7.34	7.77	8.22	8.97
	6.70	7.48	7.95	8.49	9.59
1.5	6.53	7.29	7.69	8.09	8.78
	6.68	7.42	7.87	8.35	9.31
2.0	6.42	7.18	7.61	8.09	8.68
	6.54	7.37	7.83	8.31	9.09
2.5	6.28	7.01	7.56	8.04	8.51
	6.40	7.25	7.78	8.26	8.91
3.0	6.22	6.93	7.53	8.00	8.35
	6.33	7.18	7.74	8.19	8.71
3.5	6.15	6.84	7.35	7.79	8.28
	6.25	6.99	7.52	7.99	8.71
4.0	6.17	6.81	7.21	7.74	8.55
	6.23	6.89	7.40	7.93	9.00

**Note.** The  $M_{\text{BH}}$  values below which 5%, 25%, 50%, 75%, and 95% of all the SMBH mass is included at different redshifts. The first column is the redshift, and the remaining ones list the corresponding  $\log M_{\text{BH}}$ . Within each cell, the top and bottom parts show the TNG100 and TNG300 results, respectively.

is nothing more than the corresponding TNG SMF scaled down by the scaling relation in A. E. Reines & M. Volonteri (2015). From  $z = 4$  to  $z = 1$ , the BHMF normalization increases steadily at all  $M_{\text{BH}}$ , with limited changes in the overall shape. From  $z = 1$  to  $z = 0$ , the BHMF almost remains the same, indicating that the overall BHMF has been largely built by  $z = 1$ . Although BHMFs in previous literature can agree with each other within  $\approx 1$  dex at  $z = 0$ , there have been large differences in how the BHMF is built over time. F. Shankar et al. (2009) concluded that the BHMF across the whole  $M_{\text{BH}}$  range evolved strongly between  $z = 4$  and  $z = 1$ , and its massive part at  $M_{\text{BH}} \gtrsim 10^8 M_\odot$  was mostly built up by  $z = 1$ , but there would still be large evolution by  $\approx 1$  dex for the less-massive part from  $z = 1$  to  $z = 0$ . A. Merloni & S. Heinz (2008) predicted that only the massive part of the BHMF at  $M_{\text{BH}} \gtrsim 10^{7.5} M_\odot$  evolved before  $z = 1$ , while the less-massive part began its evolution at  $z \lesssim 1$  and gradually caught up to the BHMF at  $z = 0$ . Recent work in A. Sicilia et al. (2022) predicts an evolution pattern more similar to ours, where the normalization of the BHMF steadily increases from  $z = 4$  to  $z = 1$  and then mostly locks down with little evolution down to  $z = 0$ . Cosmological simulations also predict strong BHMF evolution from  $z = 4$  to  $z = 1$  at any  $M_{\text{BH}}$  and further accumulation of massive SMBHs with  $M_{\text{BH}} \gtrsim 10^9 M_\odot$  after  $z = 1$ , but there are divergences regarding whether the BHMF below  $10^9 M_\odot$  continues to significantly evolve at  $z < 1$  (M. Habouzit et al. 2021).

We further show the  $M_{\text{BH}}$  values below which 5%, 25%, 50%, 75%, and 95% of all the SMBH mass is included as functions of  $z$  in Table 3 and Figure 8. The typical  $M_{\text{BH}} \approx 10^8 M_\odot$  at  $z = 0$ , below (above) which half of all the SMBH mass is distributed. These numbers only decrease modestly with increasing redshift because the overall BHMF shape remains similar at different redshifts (Figure 7).



**Figure 8.** The  $M_{\text{BH}}$  values below which 5%, 25%, 50%, 75%, and 95% of all the SMBH mass is included as functions of  $z$ . The black curve represents the 50% one, and the shaded region encloses 25%–75% and 5%–95% mass ranges.

### 3.3. Growth Rates in Different Channels

The accretion-driven  $M_{\text{BH}}$  growth rate,  $\dot{M}_a$ , is from F. Zou et al. (2024), and we further measure  $\dot{M}_m$  and compare it with  $\dot{M}_a$  in this section. For a given SMBH, we consider all of its immediate progenitors in its merger tree and label the most massive one as the primary one, and the mass added to this primary one is from all the remaining SMBHs. This added mass is then divided by the corresponding time interval to calculate  $\dot{M}_m$ . Note that the actual added mass depends on  $p_{\text{merge}}$  and the GW-radiation mass loss, as discussed in Section 2.4.

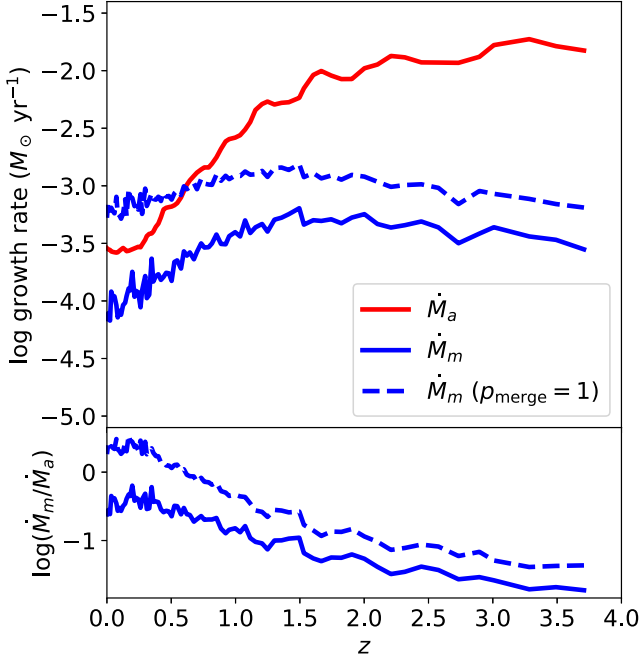
We compare  $\dot{M}_a$  and  $\dot{M}_m$  averaged over all of our TNG sources in Figure 9. Two types of  $\dot{M}_m$  with different  $p_{\text{merge}}$  are included; one is for our previously adopted  $p_{\text{merge}}$  in M. Tremmel et al. (2018b), and the other is  $p_{\text{merge}} = 1$ , which represents the upper limit for  $\dot{M}_m$ . Figure 9 indicates that  $\dot{M}_a$  significantly declines at  $z < 2$ , but  $\dot{M}_m$  shows more minor redshift evolution, with a small decline at  $z < 1$ . Therefore, the fractional contribution from the merger channel to the total SMBH growth becomes larger as redshift decreases. Figure 9 also indicates that the accretion channel always dominates over the merger channel at any redshift for the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b). In the upper-limit case where  $p_{\text{merge}} = 1$ , though, it is possible that the merger channel dominates over the accretion channel at  $z < 0.5$ . A. Ricarte & P. Natarajan (2018) adopted a constant  $p_{\text{merge}} = 0.1$ , which is more physically plausible compared to  $p_{\text{merge}} = 1$ , and the corresponding  $\dot{M}_m$  is 1 dex lower than the curve for  $p_{\text{merge}} = 1$ , leading to an  $\dot{M}_m$  similar to ours based on M. Tremmel et al. (2018b) at  $z < 0.5$  but being smaller at higher redshifts. Therefore, at least at  $z \gtrsim 0.5$ , the merger channel is negligible for the whole SMBH population, and it is likely that the merger channel also does not dominate at lower redshifts unless  $p_{\text{merge}}$  is much higher than expected.

We further measure  $\dot{M}_m$  in  $M_*$  and  $M_{\text{BH}}$  bins at different redshifts, where we bin  $M_*$  or  $M_{\text{BH}}$  with a width of 0.25 dex and only keep bins with at least 50 sources. The results are tabulated in Tables 4 and 5, and  $\dot{M}_a$  can be retrieved from Section 3.3 of F. Zou et al. (2024). Figures 10 and 11 present  $\dot{M}_m$  versus  $M_*$  and  $M_{\text{BH}}$ . We plot the 90% confidence range of  $\dot{M}_a(M_*, z)$  in F. Zou et al. (2024) and use the  $M_{\text{BH}}-M_*$  scaling relation in Section 3.1 to convert  $\dot{M}_a$  as a function of  $(M_*, z)$  to

**Table 4**  
Merger-driven Growth Rates in  $M_*$  Bins

$\log M_*$	$z = 0$	$z = 0.5$	$z = 1$	$z = 1.5$	$z = 2$	$z = 2.5$	$z = 3$
9.50–9.75	−5.62 (−5.72) −6.25 (−5.90)	−5.77 (−5.63) −6.24 (−5.87)	−5.90 (−5.62) −6.14 (−5.84)	−6.59 (−6.40) −6.22 (−5.86)	−6.40 (−6.05) −6.27 (−6.06)	−6.80 (−6.19) −6.47 (−6.20)	−7.09 (−6.74) −6.87 (−6.45)
9.75–10.00	−5.44 (−5.42) −6.08 (−5.41)	−5.63 (−5.18) −5.58 (−5.20)	−5.39 (−4.90) −5.39 (−5.04)	−5.42 (−4.95) −5.29 (−5.00)	−5.42 (−5.07) −5.27 (−4.90)	−5.45 (−5.02) −5.32 (−5.05)	−5.67 (−5.07) −5.36 (−5.25)
10.00–10.25	−5.55 (−4.77) −5.47 (−4.95)	−5.14 (−4.56) −5.10 (−4.59)	−4.78 (−4.28) −4.80 (−4.31)	−4.60 (−4.06) −4.59 (−4.21)	−4.57 (−4.12) −4.57 (−4.18)	−4.59 (−4.17) −4.52 (−4.19)	−4.57 (−4.15) −4.52 (−4.23)
10.25–10.50	−5.20 (−4.52) −5.12 (−4.39)	−4.87 (−4.13) −4.65 (−4.07)	−4.38 (−3.73) −4.27 (−3.75)	−4.14 (−3.59) −4.04 (−3.52)	−4.12 (−3.56) −4.00 (−3.51)	−4.02 (−3.54) −4.01 (−3.52)	−3.90 (−3.51) −4.00 (−3.50)
10.50–10.75	−4.99 (−4.02) −4.63 (−3.92)	−3.91 (−3.43) −4.16 (−3.60)	−3.56 (−3.08) −3.75 (−3.26)	−3.39 (−2.91) −3.51 (−3.02)	−3.33 (−2.86) −3.45 (−2.97)	−3.28 (−2.75) −3.44 (−2.98)	−3.17 (−2.70) −3.40 (−2.95)
10.75–11.00	−3.68 (−3.05) −3.85 (−3.16)	−3.23 (−2.64) −3.40 (−2.85)	−2.77 (−2.27) −2.98 (−2.52)	−2.60 (−2.19) −2.68 (−2.31)	−2.37 (−2.07) −2.70 (−2.35)	−2.63 (−2.20) −2.72 (−2.39)	−2.59 (−2.15) −2.72 (−2.40)
11.00–11.25	−3.14 (−2.37) −3.21 (−2.51)	−2.82 (−2.10) −2.69 (−2.17)	−2.43 (−1.89) −2.26 (−1.84)	−2.01 (−1.65) −2.00 (−1.65)	−2.04 (−1.74) −2.03 (−1.71)	−2.26 (−1.93) −2.07 (−1.79)	...
11.25–11.50	−1.73 (−1.77) −2.83 (−2.01)	−2.47 (−1.74) −2.41 (−1.72)	−1.92 (−1.40) −1.96 (−1.45)	...	...	...	...
11.50–11.75	−2.55 (−1.39) −2.50 (−1.61)	...	...	...	...	...	...
11.75–12.00	...	...	...	...	...	...	...
	−2.98 (−1.41)	−1.85 (−1.05)	−1.45 (−0.80)	−1.30 (−0.69)	...	...	...

**Note.** The first column lists the  $\log M_*$  range of each bin, and the other columns show  $\log \dot{M}_m$  in  $M_\odot \text{ yr}^{-1}$  at the corresponding bins. Within each cell, the top and bottom parts show the TNG100 and TNG300 results, respectively. The values in the parentheses correspond to  $p_{\text{merge}} = 1$  and are thus upper limits.



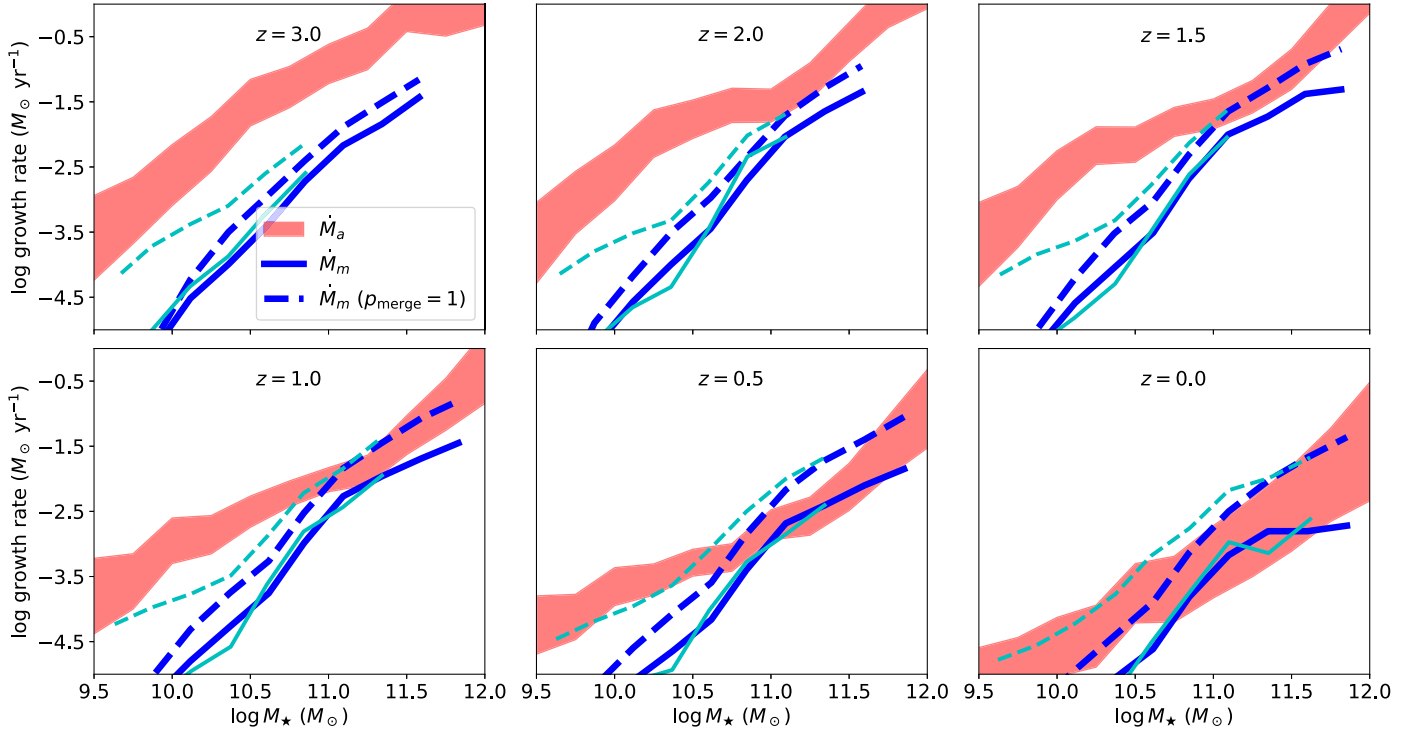
**Figure 9.** Top: the mean growth rates in the accretion channel (red) and the merger channel (blue) for our whole sample. The solid and dashed merger-channel curves correspond to the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) and  $p_{\text{merge}} = 1$ , respectively, and the latter represents the upper limit of  $\dot{M}_m$ . Bottom:  $\log(\dot{M}_m/\dot{M}_a)$  vs.  $z$ . The visually stronger fluctuations at low redshifts are due to the decreasing redshift interval length between snapshots with the same time interval.

a function of  $(M_{\text{BH}}, z)$  in Figure 11. To suppress  $\dot{M}_m$  fluctuations, we directly calculate the expected merged  $M_{\text{BH}}$  for each galaxy merger event as  $p_{\text{merge}} M_{\text{BH}}^{\text{secondary}}$ , where  $M_{\text{BH}}^{\text{secondary}}$  is the mass of the secondary SMBH, to avoid introducing additional uncertainties from the random determination of whether two SMBHs would eventually merge. We

also calculate  $\dot{M}_m$  within  $\pm 3$  snapshots (i.e., spanning 0.9 Gyr) and adopt their average value as  $\dot{M}_m$  in each bin (i.e., we use *sliding window* smoothing). In Figure 11, we also explicitly mark the predicted transitioning  $M_{\text{BH}}$  where  $\dot{M}_a = \dot{M}_m$  from F. Pacucci & A. Loeb (2020) for comparison. The figures show that both  $\dot{M}_a$  and  $\dot{M}_m$  increase with  $M_*$  and  $M_{\text{BH}}$ , and  $\dot{M}_m$  gradually approaches  $\dot{M}_a$  with rising mass and decreasing redshift. The solid blue  $\dot{M}_m$  curves begin to overlap with the  $\dot{M}_a$  shaded regions at  $z < 1$  when  $M_* \gtrsim 10^{10.7} - 10^{11.2} M_\odot$  or  $M_{\text{BH}} \gtrsim 10^{7.5} - 10^{8.5} M_\odot$ , where the exact transitioning masses move to lower masses with decreasing redshift. If we set  $p_{\text{merge}} = 1$ ,  $\dot{M}_m$  would become comparable with  $\dot{M}_a$  earlier at  $z = 1.5$  at the high-mass end ( $M_* \gtrsim 10^{11} M_\odot$  and  $M_{\text{BH}} \gtrsim 10^{8.5} M_\odot$ ), and the transitioning mass at lower redshift would decrease by  $\approx 0.3$  dex. Under this case, it is even possible that  $\dot{M}_m(z = 0)$  is comparable to  $\dot{M}_a(z = 0)$  for the whole mass range we are considering, although this is partly due to the elevated statistical uncertainties of  $\dot{M}_a(z = 0)$ .

In the analyses above, we did not consider merged masses from MBHs in dwarf galaxies, which could cause underestimation of  $\dot{M}_m$ . We first estimate our best-guess contributions from MBHs in dwarf galaxies by extending the MBH seeding down to  $M_* = 10^8 M_\odot$ , extrapolating the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) down to the dwarf-galaxy regime, and setting  $\dot{M}_a(M_* < 10^{9.5} M_\odot) = \dot{M}_a(M_* = 10^{9.5} M_\odot) \times M_*/(10^{9.5} M_\odot)$ . When seeding MBHs into dwarf galaxies, we also randomly determine if the galaxy contains an MBH seed with a Bernoulli realization, where the success probability is set to the MBH-occupation fraction in Equation (1) of E. Gallo & A. Sesana (2019). We rerun our analyses for TNG100.<sup>16</sup> The resulting best-guess  $\dot{M}_m$  is plotted in Figures 10 and 11 and is similar to

<sup>16</sup> We present TNG100 results for this analysis because it has a better baryonic mass resolution suitable to probe the dwarf regime. Besides, our further results indicate that these added contributions from dwarf galaxies, even in the most extreme case, should have little impact on the high-mass end, and thus the analysis of TNG300 is unnecessary.



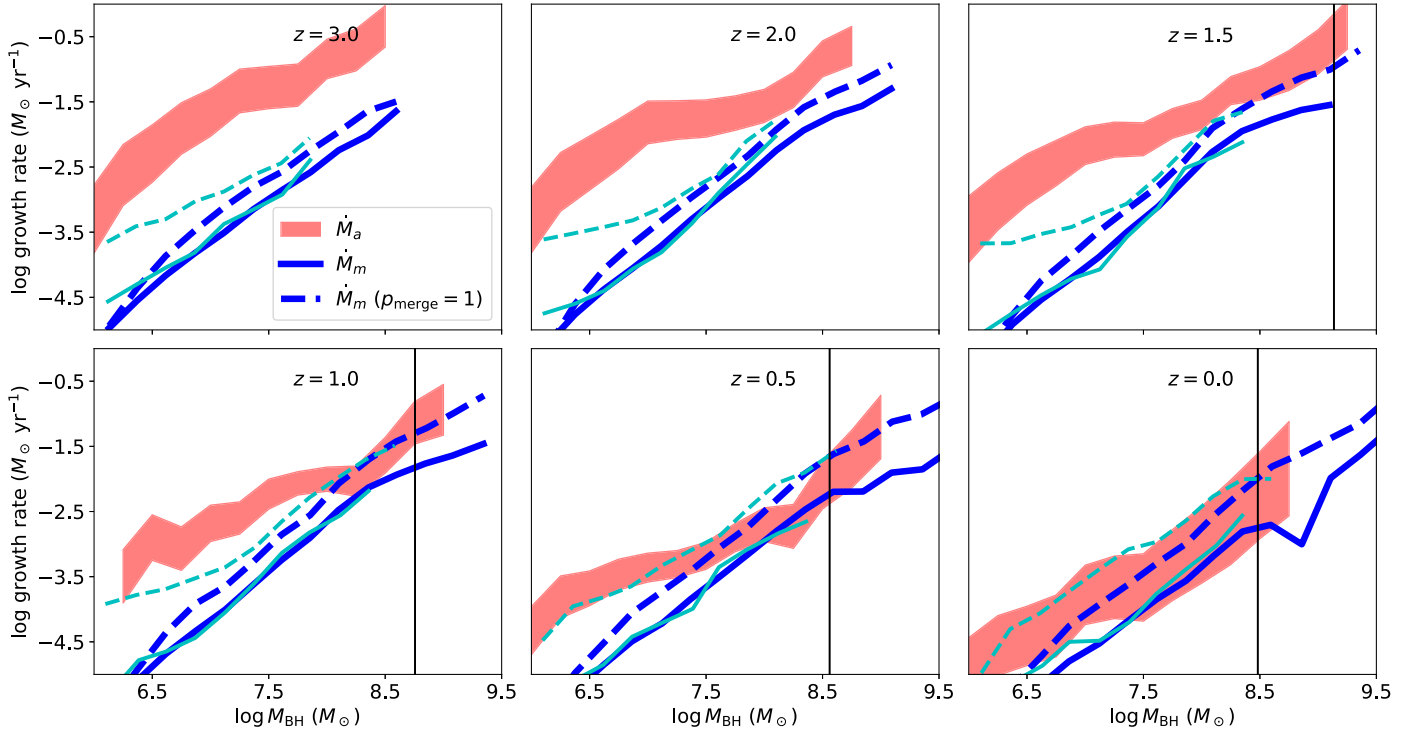
**Figure 10.** The growth rates as functions of  $M_*$  at different redshifts. The red shaded regions represent the 90% confidence range of  $\dot{M}_a$  in F. Zou et al. (2024). The blue curves represent  $\dot{M}_m$ , and the corresponding solid and dashed curves represent the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) and  $p_{\text{merge}} = 1$ , respectively. The cyan curves correspond to TNG100  $\dot{M}_m$  when considering MBHs in dwarf galaxies, where the solid and dashed ones represent our best-guess and extreme, hard-limit cases, respectively.

**Table 5**  
Merger-driven Growth Rates in  $M_{\text{BH}}$  Bins

$\log M_{\text{BH}}$	$z = 0$	$z = 0.5$	$z = 1$	$z = 1.5$	$z = 2$	$z = 2.5$	$z = 3$
6.00–6.25	−5.39 (−5.49) −5.80 (−5.73)	−5.88 (−5.63) −5.69 (−5.59)	−5.43 (−5.49) −5.55 (−5.52)	−5.38 (−5.43) −5.43 (−5.36)	−5.20 (−5.33) −5.24 (−5.31)	−5.15 (−5.14) −5.09 (−5.07)	−5.00 (−4.93) −4.97 (−4.94)
6.25–6.50	−5.38 (−5.07) −5.53 (−5.20)	−5.24 (−5.02) −5.25 (−4.98)	−5.14 (−4.91) −5.08 (−4.91)	−4.94 (−4.92) −4.92 (−4.77)	−5.00 (−4.55) −4.76 (−4.61)	−4.70 (−4.51) −4.62 (−4.45)	−4.59 (−4.33) −4.55 (−4.36)
6.50–6.75	−5.10 (−4.63) −5.20 (−4.73)	−4.90 (−4.51) −4.85 (−4.51)	−4.57 (−4.34) −4.68 (−4.36)	−4.62 (−4.33) −4.53 (−4.25)	−4.42 (−4.19) −4.38 (−4.07)	−4.21 (−3.93) −4.25 (−3.95)	−4.07 (−3.85) −4.16 (−3.86)
6.75–7.00	−5.02 (−4.25) −4.81 (−4.25)	−4.75 (−4.20) −4.49 (−4.06)	−4.48 (−3.94) −4.34 (−3.92)	−4.25 (−3.78) −4.21 (−3.86)	−4.11 (−3.75) −4.06 (−3.68)	−4.03 (−3.61) −3.93 (−3.54)	−4.11 (−3.40) −3.83 (−3.48)
7.00–7.25	−4.86 (−4.47) −4.55 (−3.90)	−4.16 (−3.69) −4.22 (−3.73)	−4.11 (−3.64) −4.00 (−3.65)	−3.93 (−3.48) −3.88 (−3.48)	−3.71 (−3.42) −3.71 (−3.32)	−3.64 (−3.24) −3.57 (−3.17)	−3.43 (−3.15) −3.51 (−3.12)
7.25–7.50	−4.46 (−3.71) −4.22 (−3.63)	−3.87 (−3.37) −3.85 (−3.42)	−3.63 (−3.27) −3.62 (−3.26)	−3.51 (−3.25) −3.49 (−3.15)	−3.30 (−2.92) −3.31 (−2.97)	−3.19 (−2.73) −3.20 (−2.87)	−3.23 (−2.65) −3.15 (−2.81)
7.50–7.75	−3.65 (−3.31) −3.80 (−3.31)	−3.66 (−3.12) −3.51 (−3.06)	−3.15 (−2.75) −3.25 (−2.86)	−3.10 (−2.75) −3.10 (−2.79)	−2.97 (−2.63) −2.97 (−2.66)	−2.91 (−2.51) −2.89 (−2.60)	−2.59 (−2.45) −2.86 (−2.57)
7.75–8.00	−3.11 (−2.72) −3.60 (−3.01)	−3.13 (−2.54) −3.15 (−2.73)	−2.74 (−2.34) −2.90 (−2.56)	−2.53 (−2.34) −2.70 (−2.40)	−2.44 (−2.23) −2.63 (−2.33)	−2.65 (−2.44) −2.58 (−2.28)	−2.61 (−2.33) −2.58 (−2.25)
8.00–8.25	−2.88 (−2.33) −3.26 (−2.58)	−2.62 (−2.15) −2.81 (−2.32)	−2.50 (−2.03) −2.48 (−2.07)	−2.27 (−1.73) −2.27 (−1.90)	−2.07 (−1.73) −2.25 (−1.93)	−2.43 (−1.87) −2.26 (−1.97)	...
8.25–8.50	−2.61 (−1.87) −2.78 (−2.19)	−2.61 (−1.93) −2.47 (−1.92)	−2.52 (−1.73) −2.13 (−1.70)	−2.13 (−1.74) −1.95 (−1.60)	...	...	...
8.50–8.75	...	...	...	...	...	...	...
8.75–9.00	−2.52 (−1.72) ...	−2.20 (−1.62) ...	−1.93 (−1.43) ...	−1.77 (−1.34) ...	−1.70 (−1.34) ...	−1.72 (−1.37) ...	−1.65 (−1.49) ...
9.00–9.25	−3.07 (−1.59) ...	−2.20 (−1.43) ...	−1.76 (−1.22) ...	−1.63 (−1.13) ...	−1.57 (−1.18) ...	−1.31 (−1.09) ...	...
9.25–9.50	−2.00 (−1.44) ...	−1.90 (−1.12) ...	−1.64 (−0.99) ...	−1.54 (−1.01) ...	−1.30 (−0.93) ...	...	...
9.50–9.75	−1.63 (−1.22) ...	−1.85 (−1.01) ...	−1.46 (−0.72) ...	...	...	...	...
	−1.23 (−1.01) ...	−1.55 (−0.77) ...	...	...	...	...	...

**Note.** Same as Table 4, but for  $\dot{M}_m$  in  $M_{\text{BH}}$  bins.





**Figure 11.** Same as Figure 10, but for the growth rates as functions of  $M_{\text{BH}}$ , where we use our  $M_{\text{BH}}-M_*$  scaling relation to convert  $M_{\text{BH}}$  to  $M_*$ . The black vertical lines are the transitioning  $M_{\text{BH}}$  predicted in F. Pacucci & A. Loeb (2020) when  $\dot{M}_a = \dot{M}_m$  at the corresponding redshifts.

the original  $\dot{M}_m$ , indicating that the contributions from dwarf galaxies may be small. However, the above best-guess contribution from dwarf galaxies may not be reliable due to the highly limited understanding of MBHs in dwarf galaxies, as outlined in Section 2.2, and we thus also provide hard upper limits. For this purpose, we set all the relevant factors to extreme values, i.e.,  $\dot{M}_a(M_* < 10^{9.5} M_\odot) = \dot{M}_a(M_* = 10^{9.5} M_\odot)$ ,  $p_{\text{merge}} = 1$ , and the MBH-occupation fraction in dwarf galaxies being 100%. The  $\dot{M}_m$  in this extreme, hard-limit case is also presented in Figures 10 and 11, and it elevates the original  $\dot{M}_m(p_{\text{merge}} = 1)$  upper limit primarily at the low-mass end ( $M_{\text{BH}} \lesssim 10^{10.5} M_\odot$  and  $M_{\text{BH}} \lesssim 10^{7.5} M_\odot$ ). Since the low-mass  $\dot{M}_m$  at  $z \gtrsim 1$  is far smaller than  $\dot{M}_a$ , this further elevated  $\dot{M}_m$  is still below  $\dot{M}_a$  and nearly does not change the overall SMBH growth at  $z \gtrsim 1$ . However, at  $z \lesssim 0.5$ , the elevation of  $\dot{M}_m$  becomes more important and may cause  $\dot{M}_m$  to dominate over  $\dot{M}_a$  at  $z = 0$ . Therefore, besides the regulation of  $p_{\text{merge}}$ , there is additional room for  $\dot{M}_m$  to reach higher values and become more important in the overall SMBH growth due to the possible contributions from MBHs in dwarf galaxies. This effect is negligible in our best-guess case, but, in principle, it is likely to cause more elevations in  $\dot{M}_m$ . In our extreme case, this effect matters mainly in the low-mass and low- $z$  regime. It should also be noted that this low-mass regime retains a high fraction of memory about our input seeds (Section 3.4) and is thus sensitive to the seeding.

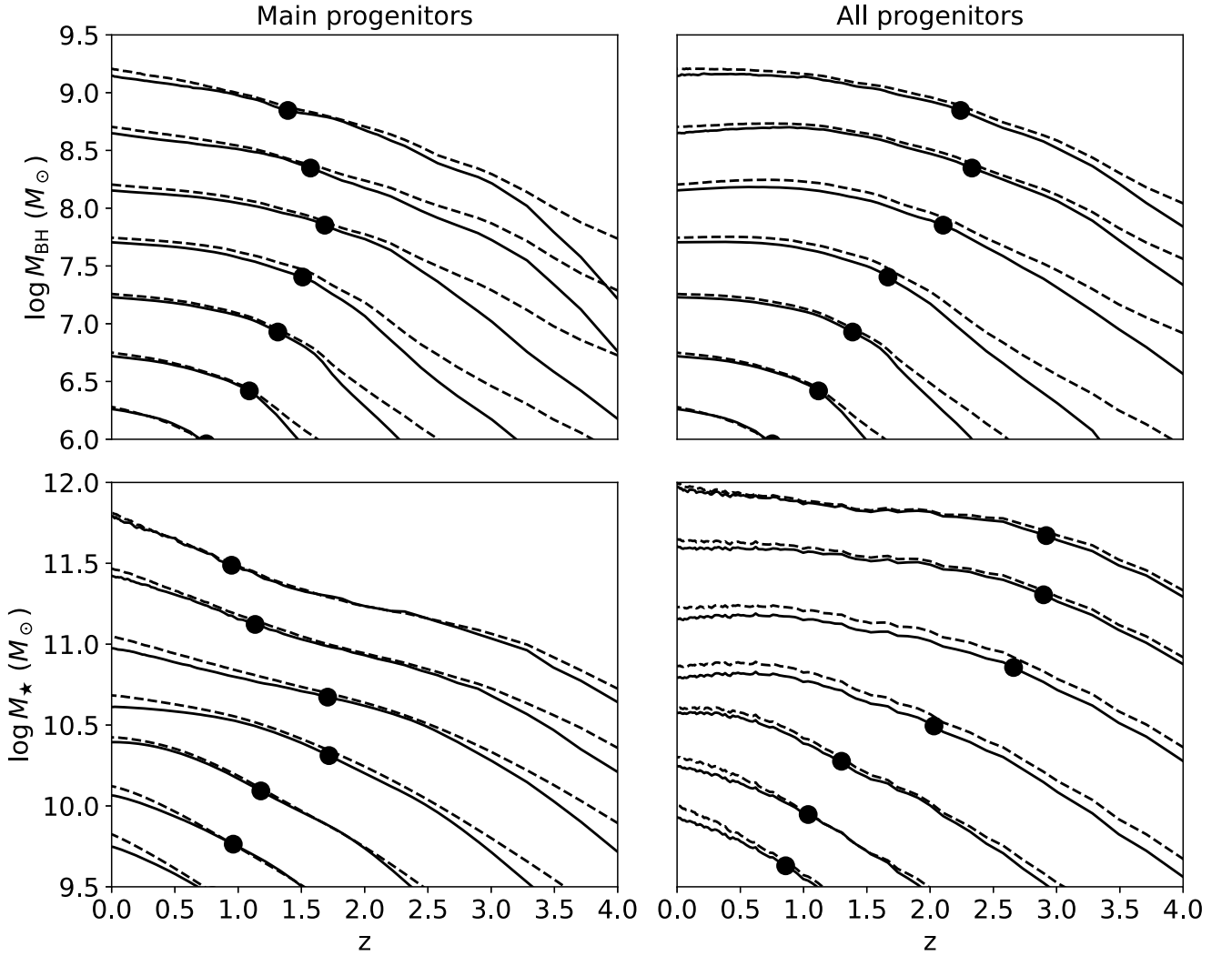
Overall, the accretion channel almost always dominates at  $z > 1$ , where most of the growth happens (see Figure 7). At lower redshift, the merger-driven growth may become comparable with the accretion-driven growth, particularly at the high-mass end with  $M_* \gtrsim 10^{11} M_\odot$ . The transitioning mass can be pushed down further, depending upon  $p_{\text{merge}}$  and contributions from MBHs in dwarf galaxies. This finding, however, differs from the original TNG-simulated results, where the TNG SMBH growth is instead dominated by

mergers for galaxies with  $M_* \gtrsim 10^{10.5} M_\odot$  or  $M_{\text{BH}} \gtrsim 10^{8.5} M_\odot$  at  $z < 1-3$  (R. Weinberger et al. 2018). The main reason is that the TNG SMBH accretion in galaxies with  $M_* \gtrsim 10^{10.5} M_\odot$  tends to be strongly suppressed due to very efficient feedback in TNG simulations, which can also be seen in Figure 5 where the TNG  $\dot{M}_a$  does not increase with  $M_*$  at  $M_* \gtrsim 10^{10.5} M_\odot$ . Since the TNG  $\dot{M}_a$  is significantly below the observational measurements for massive galaxies at  $z \lesssim 2$  (Figure 5), we argue that the accretion-driven growth should have more contribution than inferred from the original TNG simulations.

An observable metric that might indicate the relative contributions from different channels is the SMBH spin. Growth through coherent accretion (if applicable) can efficiently spin up SMBHs, while chaotic mergers tend to cause low-to-intermediate spins (e.g., C. S. Reynolds 2019 and references therein). Only about two dozen SMBHs at  $z \lesssim 1$  have spin measurements (e.g., R. V. Vasudevan et al. 2016), but they reveal the tendency that SMBHs with  $M_{\text{BH}} \lesssim 10^{7.5} M_\odot$  generally have high spins, and more massive SMBHs may have lower spins (see Figure 6 in R. V. Vasudevan et al. 2016). Although large uncertainties and selection biases exist for the currently available spin measurements, this pattern is broadly consistent with our finding that SMBHs with  $M_{\text{BH}} \lesssim 10^{7.5} M_\odot$  should generally have accretion-driven growth.

### 3.4. Mass-assembly History

We probe the redshift evolution of  $M_{\text{BH}}$  in this section. A single galaxy at  $z = 0$  usually has many high-redshift progenitors, and we adopt two kinds of definitions of the evolutionary track. First, we adopt the main-progenitor branch of its merger tree, which picks out a single progenitor galaxy (if any) with the most massive history, as defined in G. De Lucia & J. Blaizot (2007), in each relevant snapshot, as the *main-progenitor* track. Galaxies on the main-progenitor branch are



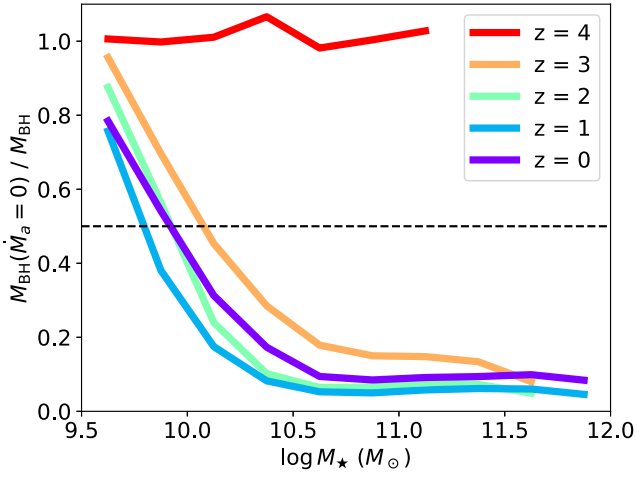
**Figure 12.** The evolutionary tracks of  $M_{\text{BH}}$  (top) and  $M_{\star}$  (bottom) for the main progenitors’ masses (left) and all the progenitors’ total masses (right), where sources are binned with an  $M_{\text{BH}}(z=0)$  width of 0.5 dex. The solid and dashed curves are  $\log[\text{median}(M_{\text{BH}})]$  ( $\log[\text{median}(M_{\star})]$ ) and  $\log[\text{mean}(M_{\text{BH}})]$  ( $\log[\text{mean}(M_{\star})]$ ) tracks, respectively. The large points on the curves represent when half of the masses are assembled.

usually but not necessarily the most massive ones at the corresponding snapshots. Second, we adopt the total mass of all the progenitors at each snapshot as the *all-progenitor* track. As we will show below, different definitions can lead to different mass-assembly behaviors.

We bin our sources with an  $M_{\text{BH}}(z=0)$  width of 0.5 dex and plot their main-progenitor and all-progenitor evolutionary tracks of  $M_{\text{BH}}$  and  $M_{\star}$  in Figure 12. We plot both  $\log[\text{median}(M_{\text{BH}})]$  ( $\log[\text{median}(M_{\star})]$ ) and  $\log[\text{mean}(M_{\text{BH}})]$  ( $\log[\text{mean}(M_{\star})]$ ) tracks. These two types of definitions can lead to different main-progenitor tracks at early stages because the mass distribution is broad and asymmetrical—for example, many galaxies may not even have SMBHs seeded yet (i.e.,  $M_{\text{BH}}=0$ ) as their  $M_{\star}$  values have not reached our threshold of  $10^{9.5} M_{\odot}$ . However, the late-stage evolutionary tracks become largely insensitive to the definition. The figure indicates that half of  $M_{\text{BH}}$  is generally assembled by  $z \approx 0.8$ – $2$ , consistent with the findings in F. Shankar et al. (2020b). Besides, smaller SMBHs generally reach their half  $M_{\text{BH}}$  at lower redshifts for the all-progenitor tracks, which is referred to as the *cosmic downsizing*, and their host galaxies have the same behavior. Regarding the main-progenitor tracks, massive galaxies with

$M_{\star}(z=0) \gtrsim 10^{11} M_{\odot}$  and their SMBHs instead show an *upsizing* phenomenon. This difference between the main-progenitor and all-progenitor tracks is inherent in the hierarchical formation, as discussed in detail in Sections 3 and 4 of E. Neistein et al. (2006).

Figure 12 indicates that the final  $M_{\text{BH}}$  is generally several orders of magnitude higher than the initially seeded  $M_{\text{BH}}$ , and thus the final  $M_{\text{BH}}$  is dominated by the accreted mass at  $z < 4$ . To further assess this point quantitatively, we rerun our analyses in Section 2 but set  $\dot{M}_a = 0$ . The ratio between the resulting  $M_{\text{BH}}(\dot{M}_a = 0)$  and our original  $M_{\text{BH}}$  represents the memory the SMBH records about its SMBH seed, where a smaller ratio indicates that more memory is lost. Note that  $M_{\text{BH}}$  can be contributed by multiple progenitors, and thus the seeding cannot be represented by a single SMBH seed; instead, we should follow the whole merger tree and use  $M_{\text{BH}}(\dot{M}_a = 0)$  to quantify the cumulative contribution from all the relevant SMBH seeds to the final  $M_{\text{BH}}$ . We plot  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}}$  as a function of  $M_{\star}$  at different redshifts in Figure 13. The  $z=4$  curve, by construction, is roughly unity. The  $z=3$  curve rapidly decreases with  $M_{\star}$  and is below  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}} = 0.5$  when  $M_{\star} > 10^{10.1} M_{\odot}$ . The transitioning  $M_{\star}$  corresponding



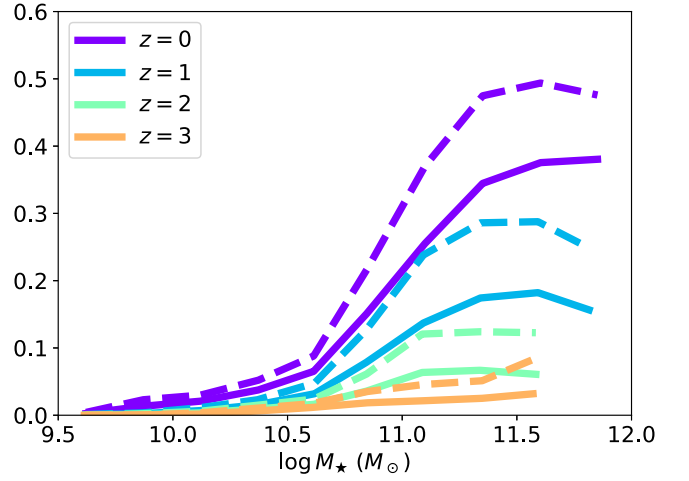
**Figure 13.** Median  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}}$  vs.  $M_*$  at different redshifts. The horizontal black dashed line represents  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}} = 0.5$ .

to  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}} = 0.5$  only slightly evolves with redshift after  $z = 3$ , partly because we must keep seeding SMBHs into newly formed massive galaxies. It is also worth noting that  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}}$  is generally above 0.5 when  $M_* < 10^{10} M_\odot$ , indicating that our results for these low-mass galaxies are largely determined by the  $M_{\text{BH}}$  seeding choice. This phenomenon has an important implication that only MBHs in low-mass galaxies with  $M_* < 10^{10} M_\odot$  record remnant information about their seeds, which has served as one pillar of the science related to MBHs in low-mass galaxies (e.g., J. E. Greene et al. 2020 and references therein). Overall, Figure 13 indicates that the seeding memory is small when  $M_* \gtrsim 10^{10} M_\odot$  and  $z \lesssim 3$ , and this is the regime where we regard this work to provide genuine physical insights into the SMBH population. We do note, though, the seeding memory would be elevated if we adopt a much higher seeding scaling relation, and Section 3.7 will discuss this in more detail.

### 3.5. Wandering SMBHs

Given the non-unity  $p_{\text{merge}}$ , some SMBHs cannot reach the final coalescence after galaxy mergers and thus become long-lived wanderers. The wandering population may constitute a non-negligible fraction of the total mass budget, and thus the classical Soltan argument needs corrections to account for wandering SMBHs (A. Kulier et al. 2015). Besides, wandering SMBHs may have important observational implications and can sometimes shine as hyper-luminous X-ray sources, dual AGNs, or off-nucleus tidal-disruption events (e.g., A. Ricarte et al. 2021a). This population is also under active theoretical research, and we briefly present our predictions based on our simplified  $p_{\text{merge}}$  method in this section.

We first show the fraction of  $M_{\text{BH}}$  locked in wandering SMBHs among all the SMBHs (i.e., central + wandering) in Figure 14, where we present both the results based on the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) and  $p_{\text{merge}} = 0$ , with the latter representing upper limits. The fraction increases with  $M_*$  and decreases with redshift. It is generally small ( $< 5\%$ ) when  $M_* < 10^{10.5} M_\odot$ , but it increases to over 30% for massive galaxies with  $M_* = 10^{11.5} M_\odot$  at  $z = 0$ . Globally, 23.4% ( $< 31.3\%$ ), 9.1% ( $< 14.9\%$ ), 3.6% ( $< 6.5\%$ ), and 1.6% ( $< 3.1\%$ ) of SMBH mass is in wandering SMBHs at  $z = 0, 1, 2$ , and 3, respectively, where the values in parentheses are the upper limits based on  $p_{\text{merge}} = 0$ . The  $z = 0$  numbers are

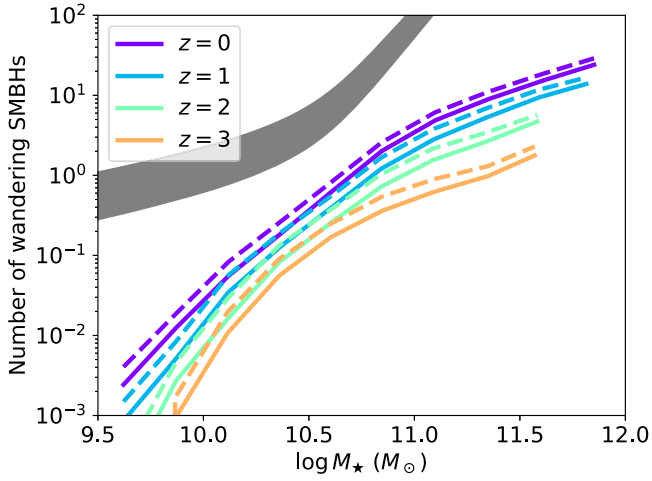


**Figure 14.** The mass fraction of wandering SMBHs among all the SMBHs vs.  $\log M_*$  at different redshifts. The solid and dashed curves correspond to the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) and  $p_{\text{merge}} = 0$ , respectively. The dashed curves represent the upper limits.

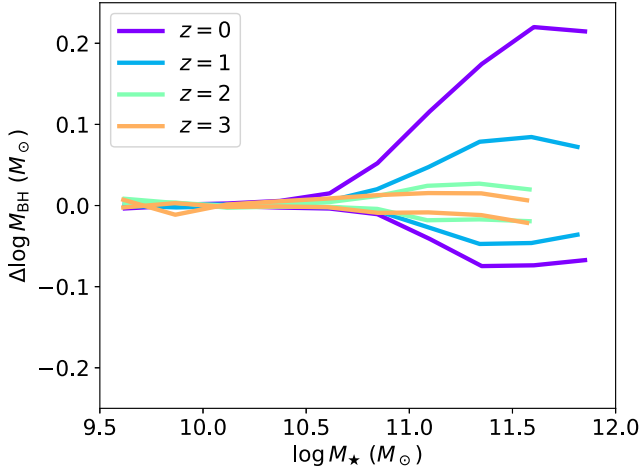
roughly consistent with previous works in an order-of-magnitude sense. A. Kulier et al. (2015) predicted the wandering mass fraction at  $z = 0$  to be 11.2% on average and strongly increasing with rising  $M_*$  and cosmic density (i.e., void versus cluster). A. Ricarte et al. (2021b) predicted the fraction to be  $\approx 10\%$ , however, with little dependence on  $M_*$ . Our value is higher possibly because our central SMBH growth is lower than in previous works (see Section 3.1). Although the mass fraction of wandering SMBHs is non-negligible, it is generally challenging to find them through direct observations because they mostly remain dormant and can hardly grow through either accretion or mergers, but observational searches for them are still possible, as predicted in, e.g., M. Guo et al. (2020) and A. Ricarte et al. (2021a).

There are some caveats worth noting. First, wandering SMBHs primarily originate from low-mass galaxies because low-mass galaxies have a higher number density than for massive galaxies, and minor mergers generally have small  $p_{\text{merge}}$ . Therefore, most of these wandering SMBHs have  $M_{\text{BH}}$  close to our initially seeded  $M_{\text{BH}}$ , and thus the wandering mass fraction is sensitive to our initial condition of SMBH seeds. Second, we neglected those wandering MBHs originating from dwarf galaxies or formed before  $z = 4$ . T. Di Matteo et al. (2023) showed that wandering MBHs may be dominated by less-massive ones, and A. Ricarte et al. (2021b) showed that wandering MBHs may even dominate over central MBHs in terms of mass at  $z = 4$ . Therefore, our lack of accounting for these MBHs may cause large underestimation of the wandering population. To illustrate this point, we plot the average number of wandering SMBHs per galaxy versus  $M_*$  in Figure 15, which shows a strong positive correlation, as expected from the hierarchical formation of galaxies. However, our number of wandering SMBHs is smaller than for A. Ricarte et al. (2021b) by  $\approx 1-2$  dex, possibly primarily because we do not consider MBHs formed in dwarf galaxies. For example, M. Tremmel et al. (2018a) predicted that MW-mass galaxies typically have roughly 12 wandering MBHs, but our predicted value is  $< 1$  because MW-like galaxies rarely undergo mergers with galaxies with  $M_* > 10^{9.5} M_\odot$  (see the left panel of Figure 2 for an example). Nevertheless, our wandering mass fraction (not the number counts) at  $z = 0$  is roughly similar to that in





**Figure 15.** The average number of wandering SMBHs per galaxy vs.  $M_*$ . The solid and dashed curves correspond to the  $p_{\text{merge}}$  in M. Tremmel et al. (2018b) and  $p_{\text{merge}} = 0$ , respectively. The gray shaded region is the relation in A. Ricarte et al. (2021b) at  $z = 0$ , where we convert the halo mass to  $M_*$  using the scaling relation in Section 3 of B. P. Moster et al. (2013). We underestimate the number of wandering MBHs because MBHs from dwarf galaxies are not considered in our analyses.

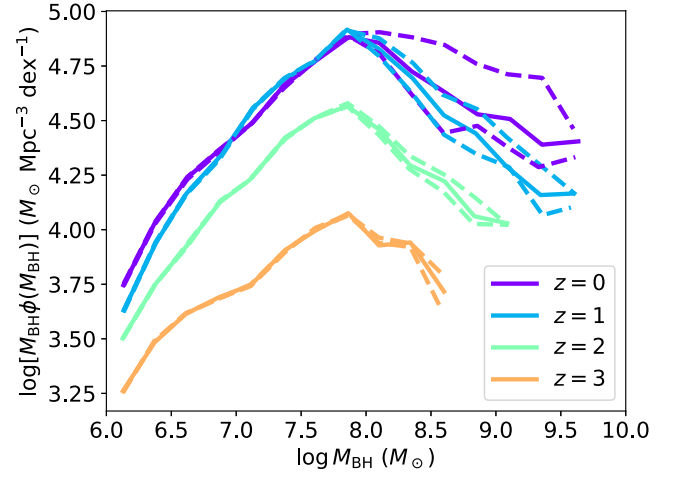


**Figure 16.** Differences in  $\log M_{\text{BH}}$ , denoted as  $\Delta \log M_{\text{BH}}$ , when changing  $p_{\text{merge}}$  to 1 and 0. At each redshift, the upper and lower curves represent the running medians of  $\log[M_{\text{BH}}(p_{\text{merge}} = 1)/M_{\text{BH}}]$  and  $\log[M_{\text{BH}}(p_{\text{merge}} = 0)/M_{\text{BH}}]$ , respectively.

A. Ricarte et al. (2021b). Lastly, we reiterate that we still have little knowledge about wandering MBHs, either from observational or theoretical perspectives. F. van Dokelaar et al. (2024) showed that the properties of wandering MBHs, such as their numbers and dynamics, identified based on cosmological simulations strongly depend on the simulation resolution and sub-grid physics modeling MBHs. Therefore, caution should be taken when interpreting the relevant results.

### 3.6. Impacts of the $p_{\text{merge}}$ Choice

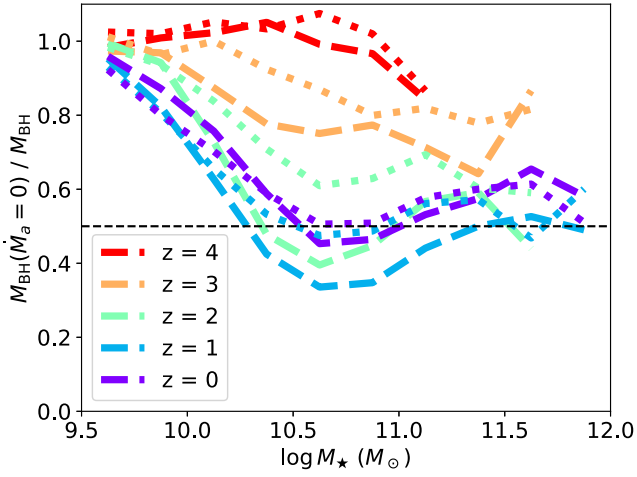
This section assesses the impact of the  $p_{\text{merge}}$  choice on our results. We choose two extreme cases of  $p_{\text{merge}} = 1$  and  $p_{\text{merge}} = 0$  and rerun the analyzes in Section 2. We plot the differences between the resulting  $\log M_{\text{BH}}$  and our original  $\log M_{\text{BH}}$  versus  $\log M_*$  in Figure 16, which also represent the deviation from the scaling relations in Figure 4. Median  $\Delta \log M_{\text{BH}}$  increases with rising  $M_*$  and decreasing  $z$ . The



**Figure 17.** The BHMf corresponding to different  $p_{\text{merge}}$  choices. At each redshift, the solid curve is the original BHMf, while the upper and lower dashed curves represent  $p_{\text{merge}} = 1$  and 0, respectively.

deviation is almost negligible ( $\lesssim 0.03$  dex) across the whole  $M_*$  range at  $z \geq 2$  and the whole  $z$  range at  $M_* \lesssim 10^{11} M_\odot$ . Difference  $\Delta \log M_{\text{BH}}$  reaches a maximum value of 0.2 dex for  $p_{\text{merge}} = 1$  and a minimum value of  $-0.1$  dex for  $p_{\text{merge}} = 0$  at  $M_* > 10^{11.5} M_\odot$  and  $z = 0$ , and such differences can hardly be the dominant element of the underlying systematic errors. We also present the BHMf corresponding to different  $p_{\text{merge}}$  in Figure 17. Similar to the findings in Figure 16, the BHMf only have noticeable differences at  $M_{\text{BH}} > 10^8 M_\odot$  and  $z \leq 1$ , where the maximum deviations are 0.3 dex and  $-0.2$  dex for  $p_{\text{merge}} = 1$  and 0 at  $z = 0$ , respectively. Integrating the BHMf, we obtain  $\rho_{\text{BH}} = 1.4 \times 10^5 M_\odot \text{Mpc}^{-3}$  and  $1.9 \times 10^5 M_\odot \text{Mpc}^{-3}$  for  $p_{\text{merge}} = 0$  and 1, respectively, which are similar to our original TNG300  $\rho_{\text{BH}}$  ( $1.5 \times 10^5 M_\odot \text{Mpc}^{-3}$ ; Section 3.2). Therefore, although some differences in  $M_{\text{BH}}$  exist for massive galaxies at  $z \leq 1$ ,  $p_{\text{merge}}$  generally does not have major impacts on  $M_{\text{BH}}$  and the relevant quantities such as the BHMf (Section 3.2) and the  $M_{\text{BH}}-M_*$  relation (Section 3.1). This is expected because Section 3.3 showed that the merger channel generally contributes little to the overall SMBH growth compared to the accretion channel, with possible exceptions at low redshifts, and  $p_{\text{merge}}$  only regulates the strength of the merger channel. In Section 3.3, we also tried elevating  $\dot{M}_m$  upper limits by considering MBHs in dwarf galaxies and showed that the resulting  $\dot{M}_m$  only differs from  $\dot{M}_m(p_{\text{merge}} = 1)$  at the low-mass end where  $\dot{M}_a$  usually dominates. Therefore, merged MBHs in dwarf galaxies almost do not impact  $M_{\text{BH}}$  (not explicitly plotted here).

However,  $p_{\text{merge}}$  plays an important role in results that are directly relevant to mergers. As has already been shown in Section 3.3,  $\dot{M}_m$  strongly depends on  $p_{\text{merge}}$ , and thus the exact transitioning  $M_{\text{BH}}$  and  $M_*$  values between the accretion-dominated ( $\dot{M}_a > \dot{M}_m$ ) and merger-dominated ( $\dot{M}_a < \dot{M}_m$ ) regimes depend upon  $p_{\text{merge}}$ . Our results on wandering SMBHs in Section 3.5 also depend upon  $p_{\text{merge}}$ . It should be noted that research on SMBH mergers and their implications is actively advancing and deserves dedicated further work (e.g., E. Barausse & A. Lapi 2021 and references therein). This work, however, does not aim to address  $p_{\text{merge}}$  but instead focuses on the overall central SMBH population, which is largely robust against the choice of  $p_{\text{merge}}$ .



**Figure 18.** Similar to Figure 13, but for different seeding choices. The dashed and dotted curves are for the high-seed and zevo-seed cases, respectively.

### 3.7. Impacts of the Seeding Choice

This section assesses the impacts of the initial seeding choice on our results. As mentioned in Section 2.2, SMBHs are initially seeded following the AGN-based  $M_{\text{BH}}-M_*$  scaling relation in A. E. Reines & M. Volonteri (2015), which is lower than those based on dynamically measured  $M_{\text{BH}}$  by over 1 order of magnitude. We thus try adopting the latter one in A. E. Reines & M. Volonteri (2015),  $\log(M_{\text{BH}}/M_\odot) = 8.95 + 1.40[\log(M_*/M_\odot) - 11]$ , as the seeding relation to examine how our results will change accordingly. We name it the *high-seed* case in this subsection. We also try applying a redshift-dependent factor in Pacucci & Loeb (2024, see their Figure 3) to scale up our original seeding relation, named the *zevo-seed* case (*zevo* is short for *evolution with redshift*) in this subsection. In this case, the seeding relation would be elevated by 0.6, 1.0, 1.3, and 1.5 dex at  $z = 1, 2, 3$ , and 4, respectively, compared to the local relation. Note that this redshift-dependent factor was not designed for  $z < 4$  as some of the assumptions in F. Pacucci & A. Loeb (2024) would fail for non-early universe, and it is also not directly derived from observations. We solely use it to illustrate how a strongly redshift-dependent seeding relation can impact our results. We rerun our analyses with these two seeding choices while keeping all of our other settings the same as in Section 2.

Before we present further scientific results, we first show memories about the seed masses,  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}}$ , in Figure 18. Unlike our original case in Figure 13, both the new high-seed and zevo-seed results retain  $\gtrsim 50\%$  of the memories about their seed masses, even in the high-mass regime, because the new seeds are much more massive than our previous ones. The fact that the median  $M_{\text{BH}}(\dot{M}_a = 0)/M_{\text{BH}} \approx 0.5$  at  $z < 2$  in the high-mass regime indicates that the new seed masses are comparable to the accreted masses for massive galaxies. Since our original  $M_{\text{BH}}$  in the high-mass regime is primarily from the accreted masses (see Figure 13), increasing seed masses to the same level of the accreted masses would cause our  $M_{\text{BH}}$  in the high-mass regime to roughly double (i.e., increase by  $\approx 0.3$  dex).

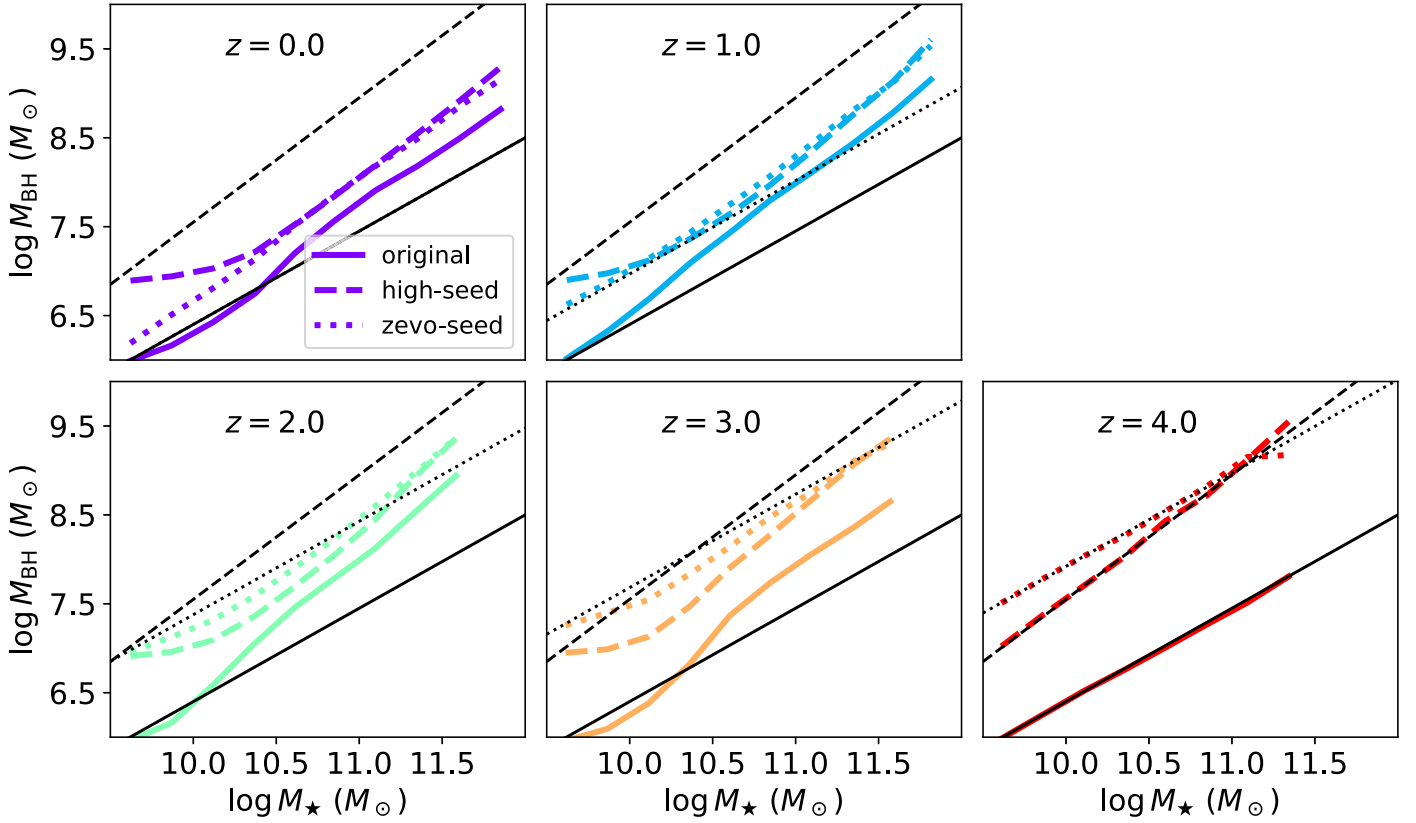
With this conclusion in mind, we compare the resulting  $M_{\text{BH}}-M_*$  scaling relations under different seeding choices in Figure 19. In the high-seed case, our recovered scaling relation at  $M_* \gtrsim 10^{10.5} M_\odot$  is higher than our originally recovered one

by over 1 dex at  $z = 4$ ,  $\approx 0.7$  dex at  $z = 3$ , and  $\approx 0.3$  dex at  $z \leq 2$ . At the low-mass end where the seed mass becomes increasingly dominant with decreasing  $M_*$ , our recovered scaling relation flattens to connect to the input seeding relation in the smallest mass bin. Overall, as the universe ages, the normalization of our recovered high-seed scaling relation keeps decreasing and becomes notably lower than the input one at  $z = 0$ . This is because our accreted masses are unable to maintain this high normalization. However, this redshift dependence contradicts observations, where the scaling-relation normalization does not show apparent evolution with redshift (e.g., Figure 10 in J. I. H. Li et al. 2023).

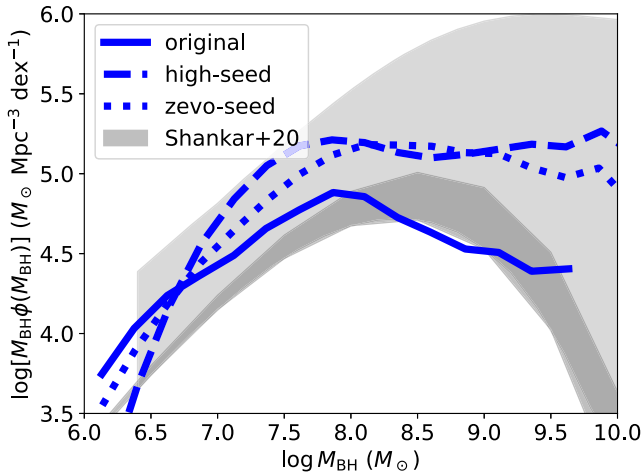
In the zevo-seed case, our recovered scaling relation is similar to the high-seed one at  $M_* \gtrsim 10^{10.5} M_\odot$  but does not become flat at the low-mass end. The seeding mainly happens at  $z > 3$  for the main progenitors of massive galaxies with  $M_*(z=0) \gtrsim 10^{10.5} M_\odot$  (see the bottom-left panel of Figure 12), while less-massive galaxies are seeded much later. Figure 19 shows that our input zevo-seed seeding relation (the dotted black lines) is similar to the high-seed one (the dashed black lines) at high redshifts and lowers down to our original one (the solid black lines) at low redshifts. Therefore, the impacts of the zevo-seed case on our results is effectively similar to the high-seed case for massive galaxies with  $M_* \gtrsim 10^{10.5} M_\odot$ . With the redshift-dependent normalization, our recovered zevo-seed scaling relation is more consistent with the input one compared to the high-seed case and also increases with redshift, similar to the input. However, it should be noted that the primary reason for the redshift dependence of our recovered relation is that the accretion at  $z < 4$  is not able to maintain the high  $M_{\text{BH}}-M_*$  normalization at  $z \approx 4$ , and SMBHs have to accumulate half of their current masses before  $z \approx 4$  for such a strongly redshift-dependent scaling relation to happen. This requires strong SMBH growth at  $z > 4$  and contradicts with previous works that the  $M_{\text{BH}}$  build-up is dominated by the  $z < 4$  epoch (e.g., F. Vito et al. 2016, 2018; F. Shankar et al. 2020b; X. Shen et al. 2020; M. Habouzit et al. 2021; H. Zhang et al. 2023). If true, such a strong SMBH growth at  $z > 4$  has to be hidden by obscuration to be missed by previous optical or X-ray surveys. Some JWST results indeed revealed stronger SMBH growth than our previous expectations (e.g., R. Maiolino et al. 2023; F. Pacucci et al. 2023), and future works are still needed to further constrain the SMBH growth in the early Universe.

With  $M_{\text{BH}}$  elevated by  $\gtrsim 0.3$  dex,  $\dot{M}_m$ , which is proportional to the merging  $M_{\text{BH}}$ , also increases by roughly the same factor. We found  $\dot{M}_m < \dot{M}_a$  at  $z \gtrsim 2$  in both the high-seed and zevo-seed cases. At lower redshift, we found that  $\dot{M}_m$  is coincidentally similar to our original  $p_{\text{merge}} = 1$  curves, with slight elevation at  $M_* < 10^{10.5} M_\odot$  and lowering at  $M_* > 10^{11.5} M_\odot$  (not explicitly plotted here). Therefore, our qualitative results in Section 3.3 remain unchanged in both the high-seed and zevo-seed cases.

However, the BHMf is more sensitive to the elevation of  $M_{\text{BH}}$ . We show the local BHMf in Figure 20. Although not plotted, the new BHMf would drop quickly at  $M_{\text{BH}} > 10^{10} M_\odot$ . The deficiency of SMBHs with  $M_{\text{BH}} < 10^7 M_\odot$  in the new BHMf is caused by our  $M_*$  cut at  $10^{9.5} M_\odot$ . Our new BHMf has more SMBHs at the massive end than our original one and are also flatter, primarily because  $M_{\text{BH}}$  becomes systematically higher. Therefore, the new BHMf becomes higher than the debiased BHMf in F. Shankar et al. (2020a). This reflects the additional uncertainty from the



**Figure 19.** Comparisons of the  $M_{\text{BH}}-M_*$  scaling relations under different seeding choices. The black straight lines are our input seeding relations at the corresponding redshifts, and the colored curves are our recovered scaling relations. The dotted black line overlaps with the solid black line at  $z = 0$  and is thus invisible. The solid curves are our original ones following the AGN-based scaling relation in A. E. Reines & M. Volonteri (2015), the dashed curves represent the high-seed case, and the dotted curves represent the zevo-seed case.



**Figure 20.** The local BHMFs under different seeding cases, as labeled in the legend. The shaded region is from F. Shankar et al. (2020a) for comparison, as explained in Figure 6.

seeding, where the BHMF may be noticeably elevated if the seed mass becomes comparable to the accreted mass.

#### 4. Summary

In this work, we use a hybrid approach to combine the observationally measured SMBH accretion in F. Zou et al. (2024) and the merger trees in the TNG simulations to probe the evolution of the SMBH population from  $z = 4$  to  $z = 0$ . Taking advantage of both observations and simulations, we aim

to construct the most realistic picture of cosmic SMBH evolution. Our main results are summarized as follows:

1. We reproduced the local  $M_{\text{BH}}-M_*$  scaling relation and showed that half of the observed scatter of this scaling relation can be accounted for by SMBH growth along diverse  $M_*$  evolutionary tracks. Our relation is lower than the original TNG-simulated relation and also dynamically measured ones but is more consistent with those based on AGNs or after corrections for the selection bias of dynamically measured SMBHs, supporting the claim that the observed dynamically measured scaling relation may be biased. We predicted that the scaling relation does not have a strong redshift evolution, at least up to  $z = 3$ , as primarily regulated by the correlation between  $\dot{M}_a$  and  $(M_*, z)$ . See Section 3.1.
2. We recovered the local BHMF and obtained a local SMBH mass density  $\rho_{\text{BH}} = 1.5 \times 10^5 M_\odot \text{Mpc}^{-3}$ . Our BHMF is lower than the original TNG-simulated ones and is more consistent with the one based on the debiased scaling relation. We predict that the BHMF steadily increases at all  $M_{\text{BH}}$  from  $z = 4$  to  $z = 1$  and largely remains frozen from  $z = 1$  to  $z = 0$ . See Section 3.2.
3. The overall SMBH growth is generally dominated by the accretion channel at  $z \gtrsim 1$ . The relative contribution from mergers becomes increasingly important with rising  $M_*$  and decreasing redshift.  $\dot{M}_m$  becomes comparable with  $\dot{M}_a$  when  $M_{\text{BH}} \gtrsim 10^8 M_\odot$  or  $M_* \gtrsim 10^{11} M_\odot$  at  $z \lesssim 1$ . The exact transitioning  $M_{\text{BH}}$  or  $M_*$  from accretion-dominated growth to merger-dominated growth depends upon  $p_{\text{merge}}$ .



and contributions from MBHs in dwarf galaxies to mergers. See Section 3.3.

4. Half of  $M_{\text{BH}}$  is accumulated by  $z \approx 0.8$ –2. Besides, SMBHs in massive galaxies with  $M_* \gtrsim 10^{10} M_\odot$  quickly lose their memories about their initial seeds at high redshift, while SMBHs in low-mass galaxies with  $M_* \lesssim 10^{10} M_\odot$  still retain information about their seed mass. See Section 3.4.
5. Around 25% of the SMBH mass budget is locked in long-lived wandering SMBHs, and the wandering mass fraction and wandering SMBH counts strongly increase with rising  $M_*$ . See Section 3.5.
6. We have examined the sensitivity of our results to the  $p_{\text{merge}}$  choice. The overall central SMBH demography is robust against  $p_{\text{merge}}$  due to the limited contributions of mergers to the overall SMBH growth, but  $p_{\text{merge}}$  regulates quantities directly relevant to mergers, such as  $\dot{M}_m$  and wandering SMBHs. See Section 3.6.
7. We also tested different seeding choices. If the dynamically measured relation is adopted, or there is a strong positive redshift evolution of the scaling-relation normalization, our recovered  $M_{\text{BH}}-M_*$  relation would be systematically elevated by  $\approx 0.3$  dex at  $z \leq 2$ , and the massive end of the BHMF would have a more noticeable difference. However, such a redshift evolution is not favored by observations (e.g., see Figure 10 in J. I. H. Li et al. 2023), and the elevation requires SMBHs to accumulate around half of their total masses by  $z = 4$ , which may also be questionable. See Section 3.7.

Overall, our results justify the power of the hybrid approach of combining observations and simulations. Further improvements will require future efforts in both directions. We are currently confined within  $M_* > 10^{9.5} M_\odot$  and  $z < 4$  by observations, and more sensitive observations, especially future deep X-ray surveys performed by the Athena, AXIS, and Lynx missions, are needed to push the  $\dot{M}_a$  measurements down to the dwarf-galaxy regime and/or into the earlier Universe. Another direction is to improve the simulations. The primary limitation, as illustrated in Figure 5 and discussed in more detail in M. Habouzit et al. (2021, 2022), is that the simulated SMBH population is inconsistent with observations, and thus the SMBH modeling in current cosmological simulations should be improved. For example, at  $z \gtrsim 2$ , the simulated  $\dot{M}_a$  is systematically higher than for observations; while at  $z \lesssim 2$ , the simulated  $\dot{M}_a$  is too high at  $M_* \lesssim 10^{11} M_\odot$  but too low at  $M_* \gtrsim 10^{11} M_\odot$ . Similarly, R. Weinberger et al. (2018) and M. Habouzit et al. (2022) showed that the TNG-simulated AGN luminosity functions are too high at high redshifts. These indicate that, for TNG, its Bondi accretion model and also its suppression of the AGN emission in massive galaxies due to the transitioning of SMBHs into the kinetic mode are overly efficient. Future simulations could use the observed AGN population properties to help calibrate their sub-grid accretion physics. Besides, future work should also try to constrain the SMBH dynamics after galaxy mergers, such as the merger probability and the time lag between SMBH mergers and galaxy mergers, which can help better characterize the merger-driven growth channel. It is particularly beneficial to construct relationships between these SMBH dynamical parameters (including their distributions) and global galaxy properties.

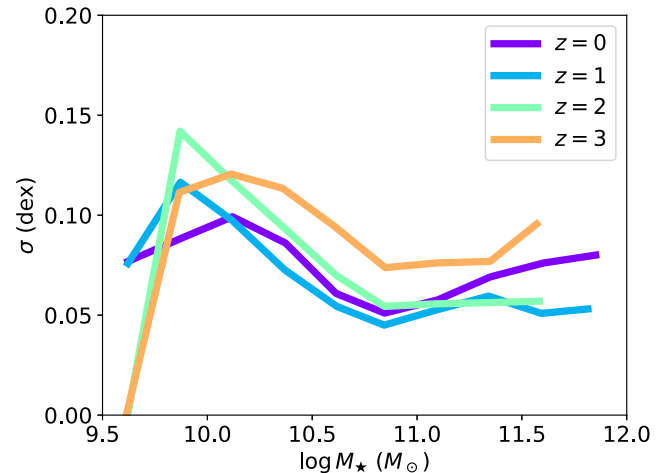
## Acknowledgments

We thank the anonymous referee for constructive suggestions and comments. We thank Lumen Boco and Andrea Lapi for providing the BHMF data. We thank Chris Byrohl, Kelly Holley-Bockelmann, Joel Leja, Dylan Nelson, Steinn Sigurdsson, and Guang Yang for helpful discussions. F.Z., W.N.B., and Z.Y. acknowledge financial support from NSF grants AST-2106990 and AST-2407089, Chandra X-ray Center grant AR4-25008X, the Penn State Eberly Endowment, and Penn State ACIS Instrument Team Contract SV4-74018 (issued by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-03060). B.L. acknowledges financial support from the National Natural Science Foundation of China grant 11991053. Y.X. acknowledges support from NSFC grants (12025303 and 12393814). The Chandra ACIS Team Guaranteed Time Observations (GTO) utilized were selected by the ACIS Instrument Principal Investigator, Gordon P. Garmire, currently of the Huntingdon Institute for X-ray Astronomy, LLC, which is under contract to the Smithsonian Astrophysical Observatory via Contract SV2-82024.

## Appendix A

### Effects of Statistical Uncertainties of the Accretion Rate

In the main text, we do not add the statistical uncertainties of the  $\dot{M}_a$  measurements in F. Zou et al. (2024) because they do not reflect any physical scatters of the SMBH population. In principle, the statistical uncertainties would approach zero with an infinite amount of data. In this Appendix, we assess the level of additional uncertainties that can be induced into  $M_{\text{BH}}$  from the  $\dot{M}_a$  measurements. We randomly draw an  $\dot{M}_a(M_*, z)$  function from the sampling results in F. Zou et al. (2024) for each source and rerun our analyzes in Section 2. We derive the additional uncertainties by comparing the new  $M_{\text{BH}}$  values with the original ones and subtracting the variance from the seeding scatter, as discussed when presenting Figure 4 in Section 3.1. We show the  $1\sigma$  uncertainty versus  $M_*$  in Figure 21. The figure shows that the additional uncertainty from the statistical uncertainties of  $\dot{M}_a$  is only  $\approx 0.05$ –0.1 dex across the whole parameter space, which should be smaller than possible systematic uncertainties. Additionally, this effect almost does not affect the median values of our results. Therefore, the



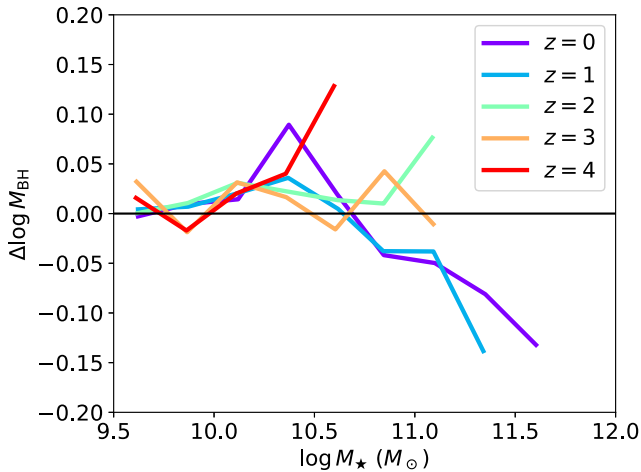
**Figure 21.** The expected additional uncertainties of  $M_{\text{BH}}$  in dex from the statistical uncertainties of  $\dot{M}_a$ , color coded by the redshift.

statistical uncertainties of  $\dot{M}_a$  are largely negligible for our  $M_{\text{BH}}$  measurements. Nevertheless, it should be noted that the elevated  $\dot{M}_a$  uncertainties at  $z \approx 0$  become more important when comparing the relative contributions of  $\dot{M}_a$  and  $\dot{M}_m$  in Section 3.3.

## Appendix B

### Comparisons between the TNG100 and TNG300 Results

We focused on TNG300 in the main text, and this Appendix compares the results based on TNG100 and TNG300. We show the difference between the TNG100 and TNG300 scaling relations in Figure 22. The figure indicates that the two scaling relations are similar at all redshifts, with differences generally less than 0.05 dex. Similarly, we have compared all of our other analyzes, and TNG100 results are also included in all of our tables for completeness. We do not explicitly plot all of them for conciseness, but the TNG100 and TNG300 results are very similar. The only more noteworthy difference is that the TNG300 SMF has a slightly heavier tail at the massive end than for TNG100, causing the mean TNG100  $M_{\text{BH}}$  to shift downward by 0.1–0.2 dex compared to TNG300. This effect can be seen more clearly in Table 3 but becomes insignificant when comparing parameters conditioned on a fixed  $M_*$ .



**Figure 22.** The difference between the median TNG100 and TNG300  $\log M_{\text{BH}}$  vs.  $M_*$  at different redshifts. The difference is generally small, indicating that TNG100 and TNG300  $M_{\text{BH}}-M_*$  scaling relations are similar.

## ORCID iDs

Fan Zou <https://orcid.org/0000-0002-4436-6923>  
W. N. Brandt <https://orcid.org/0000-0002-0167-2453>  
Elena Gallo <https://orcid.org/0000-0001-5802-6041>  
Bin Luo <https://orcid.org/0000-0002-9036-0063>  
Qingling Ni <https://orcid.org/0000-0002-8577-2717>  
Yongquan Xue <https://orcid.org/0000-0002-1935-8104>  
Zhibo Yu <https://orcid.org/0000-0002-6990-9058>

## References

Agazie, G., Anumalapudi, A., Archibald, A. M., et al. 2023, *ApJL*, **952**, L37  
Aird, J., Coil, A. L., & Georgakakis, A. 2018, *MNRAS*, **474**, 1225  
Amiri, A., Tavasoli, S., & De Zotti, G. 2019, *ApJ*, **874**, 140  
Barausse, E., Dvorkin, I., Tremmel, M., Volonteri, M., & Bonetti, M. 2020, *ApJ*, **904**, 16

Barausse, E., & Lapi, A. 2021, in *Handbook of Gravitational Wave Astronomy*, ed. C. Bambi, S. Katsanevas, & K.D. Kokkotas (Springer, Singapore), 18  
Bonetti, M., Sesana, A., Haardt, F., Barausse, E., & Colpi, M. 2019, *MNRAS*, **486**, 4044  
Brandt, W. N., & Alexander, D. M. 2015, *A&ARv*, **23**, 1  
Brandt, W. N., & Yang, G. 2022, in *Handbook of X-ray and Gamma-Ray Astrophysics*, ed. C. Bambi & A. Santangelo (Springer, Singapore), 78  
Chen, N., Ni, Y., Tremmel, M., et al. 2022, *MNRAS*, **510**, 531  
Davis, S. W., & Laor, A. 2011, *ApJ*, **728**, 98  
Davis, S. W., & Tchekhovskoy, A. 2020, *ARA&A*, **58**, 407  
De Lucia, G., & Blaizot, J. 2007, *MNRAS*, **375**, 2  
Di Matteo, T., Ni, Y., Chen, N., et al. 2023, *MNRAS*, **525**, 1479  
Dosopoulou, F., & Antonini, F. 2017, *ApJ*, **840**, 31  
Duras, F., Bongiorno, A., Ricci, F., et al. 2020, *A&A*, **636**, A73  
Edgar, R. 2004, *NewAR*, **48**, 843  
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022, *ApJL*, **930**, L15  
Fantin, N. J., Côté, P., McConnachie, A. W., et al. 2019, *ApJ*, **887**, 148  
Gallo, E., & Sesana, A. 2019, *ApJ*, **883**, L18  
Gehrels, N. 1986, *ApJ*, **303**, 336  
Genel, S., Vogelsberger, M., Springel, V., et al. 2014, *MNRAS*, **445**, 175  
Graham, A. W., & Driver, S. P. 2007, *MNRAS*, **380**, L15  
Greene, J. E., Strader, J., & Ho, L. C. 2020, *ARA&A*, **58**, 257  
Guo, M., Inayoshi, K., Michiyama, T., & Ho, L. C. 2020, *ApJ*, **901**, 39  
Habouzit, M., Li, Y., Somerville, R. S., et al. 2021, *MNRAS*, **503**, 1940  
Habouzit, M., Somerville, R. S., Li, Y., et al. 2022, *MNRAS*, **509**, 3015  
Healy, J., Lousto, C. O., & Zlochower, Y. 2014, *PhRvD*, **90**, 104004  
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, *ApJ*, **654**, 731  
Izquierdo-Villalba, D., Bonoli, S., Dotti, M., et al. 2020, *MNRAS*, **495**, 4681  
Kammoun, E. S., Miller, J. M., Koss, M., et al. 2020, *ApJ*, **901**, 161  
Kelly, B. C., & Merloni, A. 2012, *AdAst*, **2012**, 970858  
Kocevski, D. D., Finkelstein, S. L., Barro, G., et al. 2024, arXiv:2404.03576  
Kormendy, J., & Ho, L. C. 2013, *ARA&A*, **51**, 511  
Kulier, A., Ostriker, J. P., Natarajan, P., Lackner, C. N., & Cen, R. 2015, *ApJ*, **799**, 178  
Latif, M. A., & Ferrara, A. 2016, *PASA*, **33**, e051  
Lawrence, A. 2018, *NatAs*, **2**, 102  
Li, J., Silverman, J. D., Ding, X., et al. 2021a, *ApJ*, **922**, 142  
Li, J., Xue, Y., Sun, M., et al. 2019, *ApJ*, **877**, 5  
Li, J. I. H., Shen, Y., Ho, L. C., et al. 2021b, *ApJ*, **906**, 103  
Li, J. I. H., Shen, Y., Ho, L. C., et al. 2023, *ApJ*, **954**, 173  
Licquia, T. C., & Newman, J. A. 2015, *ApJ*, **806**, 96  
Liu, D., Schinnerer, E., Groves, B., et al. 2019, *ApJ*, **887**, 235  
Lodato, G., Nayakshin, S., King, A. R., & Pringle, J. E. 2009, *MNRAS*, **398**, 1392  
Lousto, C. O., Zlochower, Y., Dotti, M., & Volonteri, M. 2012, *PhRvD*, **85**, 084015  
Maiolino, R., Scholtz, J., Curtis-Lake, E., et al. 2023, arXiv:2308.01230  
Marconi, A., Risaliti, G., Gilli, R., et al. 2004, *MNRAS*, **351**, 169  
Marinacci, F., Vogelsberger, M., Pakmor, R., et al. 2018, *MNRAS*, **480**, 5113  
Merloni, A., & Heinz, S. 2008, *MNRAS*, **388**, 1011  
Miller, B. P., Gallo, E., Greene, J. E., et al. 2015, *ApJ*, **799**, 98  
Mingarelli, C. M. F. 2019, *NatAs*, **3**, 8  
Moster, B. P., Naab, T., & White, S. D. M. 2013, *MNRAS*, **428**, 3121  
Naiman, J. P., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **477**, 1206  
Neistein, E., van den Bosch, F. C., & Dekel, A. 2006, *MNRAS*, **372**, 933  
Nelson, D., Pillepich, A., Springel, V., et al. 2018, *MNRAS*, **475**, 624  
Nelson, D., Pillepich, A., Springel, V., et al. 2019a, *MNRAS*, **490**, 3234  
Nelson, D., Springel, V., Pillepich, A., et al. 2019b, *ComAC*, **6**, 2  
Ni, Q., Brandt, W. N., Yang, G., et al. 2021, *MNRAS*, **500**, 4989  
Pacucci, F., & Loeb, A. 2020, *ApJ*, **895**, 95  
Pacucci, F., & Loeb, A. 2024, *ApJ*, **964**, 154  
Pacucci, F., Nguyen, B., Carniani, S., Maiolino, R., & Fan, X. 2023, *ApJL*, **957**, L3  
Pesce, D. W., Palumbo, D. C. M., Narayan, R., et al. 2021, *ApJ*, **923**, 260  
Pillepich, A., Nelson, D., Hernquist, L., et al. 2018a, *MNRAS*, **475**, 648  
Pillepich, A., Nelson, D., Springel, V., et al. 2019, *MNRAS*, **490**, 3196  
Pillepich, A., Springel, V., Nelson, D., et al. 2018b, *MNRAS*, **473**, 4077  
Quinlan, G. D. 1996, *NewA*, **1**, 35  
Reines, A. E., & Volonteri, M. 2015, *ApJ*, **813**, 82  
Reynolds, C. S. 2019, *NatAs*, **3**, 41  
Ricarte, A., & Natarajan, P. 2018, *MNRAS*, **474**, 1995  
Ricarte, A., Tremmel, M., Natarajan, P., & Quinn, T. 2021a, *ApJL*, **916**, L18  
Ricarte, A., Tremmel, M., Natarajan, P., Zimmer, C., & Quinn, T. 2021b, *MNRAS*, **503**, 6098

- Ricci, C., Ueda, Y., Koss, M. J., et al. 2015, *ApJL*, **815**, L13
- Rodriguez-Gomez, V., Genel, S., Vogelsberger, M., et al. 2015, *MNRAS*, **449**, 49
- Rodriguez-Gomez, V., Snyder, G. F., Lotz, J. M., et al. 2019, *MNRAS*, **483**, 4140
- Sato-Polito, G., Zaldarriaga, M., & Quataert, E. 2023, arXiv:2312.06756
- Savorgnan, G. A. D., Graham, A. W., Marconi, A., & Sani, E. 2016, *ApJ*, **817**, 21
- Schmidt, M. 1963, *Natur*, **197**, 1040
- Shankar, F., Allevato, V., Bernardi, M., et al. 2020a, *NatAs*, **4**, 282
- Shankar, F., Bernardi, M., Sheth, R. K., et al. 2016, *MNRAS*, **460**, 3119
- Shankar, F., Weinberg, D. H., Marsden, C., et al. 2020b, *MNRAS*, **493**, 1500
- Shankar, F., Weinberg, D. H., & Miralda-Escudé, J. 2009, *ApJ*, **690**, 20
- Shen, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2020, *MNRAS*, **495**, 3252
- Sicilia, A., Lapi, A., Boco, L., et al. 2022, *ApJ*, **934**, 66
- Sijacki, D., Vogelsberger, M., Genel, S., et al. 2015, *MNRAS*, **452**, 575
- Small, T. A., & Blandford, R. D. 1992, *MNRAS*, **259**, 725
- Somerville, R. S., & Davé, R. 2015, *ARA&A*, **53**, 51
- Song, H., Park, C., Lietzen, H., & Einasto, M. 2016, *ApJ*, **827**, 104
- Sotan, A. 1982, *MNRAS*, **200**, 115
- Springel, V., Pakmor, R., Pillepich, A., et al. 2018, *MNRAS*, **475**, 676
- Suh, H., Civano, F., Trakhtenbrot, B., et al. 2020, *ApJ*, **889**, 32
- Sun, M., Trump, J. R., Brandt, W. N., et al. 2015, *ApJ*, **802**, 14
- Tremmel, M., Governato, F., Volonteri, M., Pontzen, A., & Quinn, T. R. 2018a, *ApJL*, **857**, L22
- Tremmel, M., Governato, F., Volonteri, M., Quinn, T. R., & Pontzen, A. 2018b, *MNRAS*, **475**, 4967
- Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, *ApJ*, **786**, 104
- van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013, *ApJL*, **771**, L35
- van Dokkelaar, F., Mayer, L., Capelo, P. R., & Tamfal, T. 2024, arXiv:2404.15404
- Varma, V., Gerosa, D., Stein, L. C., Hébert, F., & Zhang, H. 2019, *PhRvL*, **122**, 011101
- Vasiliev, E., Antonini, F., & Merritt, D. 2014, *ApJ*, **785**, 163
- Vasudevan, R. V., Fabian, A. C., Reynolds, C. S., et al. 2016, *MNRAS*, **458**, 2012
- Vito, F., Brandt, W. N., Yang, G., et al. 2018, *MNRAS*, **473**, 2378
- Vito, F., Gilli, R., Vignali, C., et al. 2016, *MNRAS*, **463**, 348
- Wang, Y., & Zhang, B. 2024, arXiv:2403.06416
- Weinberger, R., Springel, V., Hernquist, L., et al. 2017, *MNRAS*, **465**, 3291
- Weinberger, R., Springel, V., Pakmor, R., et al. 2018, *MNRAS*, **479**, 4056
- White, S. D. M., & Rees, M. J. 1978, *MNRAS*, **183**, 341
- Xiang, M., & Rix, H.-W. 2022, *Natur*, **603**, 599
- Xue, Y. Q. 2017, *NewAR*, **79**, 59
- Yan, W., Brandt, W. N., Zou, F., et al. 2023, *ApJ*, **951**, 27
- Yang, G., Brandt, W. N., Alexander, D. M., et al. 2019, *MNRAS*, **485**, 3721
- Yang, G., Brandt, W. N., Alexander, D. M., et al. 2022, *ApJ*, **940**, 146
- Yang, G., Brandt, W. N., Darvish, B., et al. 2018a, *MNRAS*, **480**, 1022
- Yang, G., Brandt, W. N., Vito, F., et al. 2018b, *MNRAS*, **475**, 1887
- Yang, G., Caputi, K. I., Papovich, C., et al. 2023, *ApJL*, **950**, L5
- Yang, G., Chen, C. T. J., Vito, F., et al. 2017, *ApJ*, **842**, 72
- Yang, G., Estrada-Carpenter, V., Papovich, C., et al. 2021, *ApJ*, **921**, 170
- Yu, Q. 2002, *MNRAS*, **331**, 935
- Zhang, H., Behroozi, P., Volonteri, M., et al. 2023, *MNRAS*, **518**, 2123
- Zou, F., Brandt, W. N., Chen, C.-T., et al. 2022, *ApJS*, **262**, 15
- Zou, F., Brandt, W. N., Ni, Q., et al. 2023, *ApJ*, **950**, 136
- Zou, F., Yu, Z., Brandt, W. N., et al. 2024, *ApJ*, **964**, 183
- Zubovas, K. 2019, *MNRAS*, **483**, 1957