

A two-phase model of galaxy formation: IV. Seeding and growing supermassive black holes in dark matter halos

Yangyao Chen,^{1,2,3*} Houjun Mo⁴ and Huiyuan Wang^{2,3}

¹*School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*

²*Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, China*

³*School of Astronomy and Space Science, University of Science and Technology of China, Hefei, Anhui 230026, China*

⁴*Department of Astronomy, University of Massachusetts, Amherst, MA 01003, USA*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We present a theoretical framework for seeding and growing supermassive black holes (SMBHs) in dark-matter halos along their assembly histories. Seeds are bred out of Pop-III stars formed during the first collapse of pristine gas in mini-halos that have reached the H₂-cooling limit, modulated by UV radiation from star formation and dynamical heating from fast halo assembly. Such breeding persists until the enrichment of the intergalactic medium (IGM) enables Pop-II stars to form. Post-seeding growth of black holes (BHs) is driven by distinct channels, starting with episodic super-Eddington accretion associated with nuclear bursts induced by global disturbances of galaxies, followed by sustained sub-Eddington accretion via capturing sub-clouds formed in self-gravitating gas clouds (SGCs) in halos of fast assembly, and ending with merger-dominated, quiescent growth. We implement the model in subhalo merger trees to build a coherent framework to follow the SMBH-galaxy-halo co-evolution over their full lives in the full structure hierarchy. BH seeds are bred with a broad mass spectrum of $M_{\text{BH}} = 10\text{--}10^5 M_{\odot}$ at $z \approx 20\text{--}30$ in mini-halos with masses of $10^5\text{--}10^8 M_{\odot}$. Nuclear bursts provide the key condition for seeds to grow into SMBHs. The $M_{\text{BH}}\text{--}M_{\star}$ relation is a multi-piece, redshift-dependent function shaped by the interplay among different growth channels. Our model predictions are broadly **consistent with existing observations**; especially, a population of BHs reminiscent of ‘little red dots’ (LRDs) discovered by JWST naturally results from the seeding and growing processes. Potential future tests of the model are discussed.

Key words: galaxies: high-redshift – galaxies: haloes – galaxies: formation – quasars: supermassive black holes

1 INTRODUCTION

The observable Universe exhibits a structure hierarchy composed of populations of objects across multiple scales (e.g. Mo et al. 2010). Galaxies, luminous stellar islands characterized by distinct luminosities and morphologies, constitute the building blocks of this hierarchy (Peacock 2003; Abazajian et al. 2003; Wright et al. 2016). On larger scales, galaxies are gravitationally clustered in groups and clusters of galaxies (Abell et al. 1989; Lumsden et al. 1992; Dalton et al. 1997; Yang et al. 2003; Van Den Bosch et al. 2003; Yang et al. 2007; Tinker 2021; Wang et al. 2020), which themselves are connected through complex structures such as filaments and sheets to form the cosmic web (Bond et al. 1996; Hahn et al. 2007; Wang et al. 2016; Zhang et al. 2025a). Within individual galaxies, the microscopic architecture, including individuals and clusters of stars and black holes, exhibits diverse properties and interacts with the host galaxy through environmental (Springel & Hernquist 2005; Dekel et al. 2009; Pimbblet 2011; Wang et al. 2018; Wise et al. 2019; Latif et al. 2022a; Wang et al. 2023; Meng et al. 2023; Puskás et al. 2025) and feedback processes (Croton et al. 2006; Bower et al. 2006; Somerville et al. 2008a; Booth & Schaye 2009; Vogelsberger et al.

2013). According to modern cosmology, these objects do not exist statically as they are observed. Instead, they are all seeded from primordial density perturbations in the early Universe, amplified and shaped by various physical processes over the long history of the Universe. With the advent of powerful telescopes, such as the James Webb Space Telescope (JWST) and Atacama Large Millimeter Array (ALMA), it is now possible to observe these objects from their infancy all the way to the present day (Finkelstein et al. 2023; Bezan-son et al. 2022; Parlanti et al. 2023; Rowland et al. 2024; Fujimoto et al. 2024; Suess et al. 2024; Naidu et al. 2024; Wang et al. 2025; Jia et al. 2025; Scholtz et al. 2025). This allows us to gather discrete evolutionary snapshots in observation to form a unified picture of structure formation. Theoretically, great efforts have been made in understanding the origin and evolution of these objects in the Λ cold dark matter (Λ CDM) paradigm (e.g. Peebles 1980; Davis et al. 1985; O’Shea & Norman 2008; Schaye et al. 2015; Stacy et al. 2016; Wang et al. 2015; Pillepich et al. 2018; Davé et al. 2019; Latif et al. 2021; Hopkins et al. 2023). This paradigm provides not only a unified cosmological background to interpret observational data, but also initial conditions, matter/energy contents and physical principles to model the evolutionary processes and to predict the structure hierarchy over the entire history of the Universe. A lot of progress has been made in modeling structure formation within this paradigm using differ-

* E-mail: yangyaochen.astro@foxmail.com

ent approaches (e.g. [Kauffmann et al. 1993](#); [Peacock & Smith 2000](#); [Springel & Hernquist 2003](#)), but many problems remain unresolved. The main challenge comes from the many components that make up the structure hierarchy, the wide range of scales involved in the processes that drive the evolution, and the complex interplay among these processes.

A key step in tackling the challenge is, therefore, to build a coherent model that can cover the whole structure hierarchy and the full evolution history within Λ CDM, so that the connection and comparison between model and observation can be made across the whole structure hierarchy and over the entire history of the Universe. In a recent paper ([Mo et al. 2024](#), hereafter [Paper-I](#)) we proposed a ‘two-phase’ scenario for galaxy formation in the context of current cosmology. The key hypothesis of this scenario is that the growth of each central galaxy can be divided into two phases according to the growth rate of its host halo: a fast phase, in which the halo accretion is fast and there is a rapid change in the gravitational potential, and a slow phase, in which the halo accretion is slow and the gravitational potential remains steady. The distinct physical processes and the results expected from them in the two phases are discussed extensively in [Paper-I](#) and modeled in other papers of this series. Here we provide a brief description that leads to the investigation presented in this paper.

- (i) **Dark matter halos.** The profile of a halo has a nearly constant concentration parameter, $c_v = c_f \approx 4$, in the fast phase, and an increasing concentration with time in the slow phase. This was found by [Zhao et al. \(2003\)](#) using N-body simulations and explained as a direct outcome of the difference in the halo accretion rate between the fast and slow phases (e.g. [Lu et al. 2006](#); [Zhao et al. 2009](#); [Klypin et al. 2016](#)). A halo in its slow phase thus has a stabilized inner core formed during the fast phase, with a radius of $R_f \equiv c_f R_s$, where R_s is the scale radius of the NFW profile ([Navarro et al. 1997](#); see also Eq. 6 below), and grows mainly by accreting mass onto the outer region.
- (ii) **Galaxies.** In the early fast phase of formation, gas can cool effectively before the halo reaches a mass of $M_v \gtrsim M_{\text{cool}} \sim 10^{12} M_\odot$, and collapses rapidly into the halo center, forming a self-gravitating gas cloud (SGC) before rotation-support becomes important. Jeans instability triggers fragmentation of the SGC, which, along with the compression by supersonic turbulence driven by the fast accretion of the halo, leads to the formation of dense sub-clouds that move almost ballistically within the galaxy. The SGC made up of such sub-clouds cannot mix angular momentum effectively to form a gaseous thin disk, but rather forms a dynamically-hot structure. Stars that form in the sub-clouds inherit the hot dynamics, forming a dynamically-hot stellar system, such as a bulge and a thick disk. Stable cold gaseous/stellar disks form and evolve only in the slow phase when the inner gravitational potential of halo is stabilized and the amount of gas is reduced by feedback processes so that rapid formation of dense sub-clouds is suppressed. To highlight the distinct dynamical hotness originated from the halo assembly, we sometimes refer to the two phase as the ‘*bulge phase*’ and the ‘*disk phase*’, respectively. As detailed in [Chen et al. \(2024\)](#), hereafter [Paper-II](#)), the condition for the formation of the SGC naturally leads to a homology relation between the size of the bulge and halo virial radius in the fast phase, $r_{\text{bulge}} \approx (1/100)R_f$, for all central galaxies, independent of halo mass and redshift, and provides an explanation for the size-mass relation of dynamically-hot galaxies observed at different redshifts.
- (iii) **Star clusters.** The density of sub-clouds can be elevated by two distinct channels, compression by supersonic turbulence and slow cooling due to low metallicity, before they fragment to form stars. Some sub-clouds can reach a ‘*supernova-free*’ density, n_{snf} , at which

the free-fall timescale (t_{ff}) of the gas falls below $t_{\text{snf}} \approx 1$ Myr, so that the gas is converted into stars before the onset of supernova feedback. Two populations of globular clusters (GCs), Pop-I and Pop-II, form from sub-clouds with density elevated by the two channels, respectively. The model and predictions related to GCs are presented in [Chen et al. \(2025a\)](#), hereafter [Paper-III](#)).

- (iv) **Supermassive black holes (SMBHs).** In the fast phase, inefficient mixing of angular momentum among sub-clouds preserves the wide distribution of the specific angular momentum (j) of sub-clouds, initially set by the turbulence. Sub-clouds with specific angular momentum lower than some threshold, j_{cap} , can then reach to the halo center to feed the central black hole (BH) of the galaxy. Here and in the following, we use BH to refer to the central black hole of a host galaxy, unless specified otherwise. The fraction of the gas captured by the BH thus depends on the mass and size of the SGC, which determine the width of the j -distribution, and on the black hole mass (M_{BH}), which determines j_{cap} . This, together with the regulation of the total gas mass of the SGC by AGN feedback, shapes the $M_{\text{BH}}-M_*$ relation for dynamically-hot galaxies observed at $z \approx 0$. In contrast, the BH growth and its feedback effects become much weaker after the galaxy transits into the slow phase where the formation of a rotation-supported disk prevents gas from sinking towards the center. This produces an offset in the $M_{\text{BH}}-M_*$ relation for dynamically-cold galaxies relative to that for dynamically-hot galaxies. Galaxy mergers play an increasingly important role in the growth of both M_* and M_{BH} for galaxies residing in more massive halos, because of the reduced amount of cold gas fraction for in-situ growth and because of the increase of merging rate with halo mass.

A critical condition for the build-up of the $M_{\text{BH}}-M_*$ relation is the effective regulation by the feedback from the accreting BH. However, this can happen only when the BH has grown massive enough so that the radiative efficiency of the associated AGN, which is $\propto M_{\text{BH}}$, is sufficient to affect gas cooling and collapse. One way to achieve this in modeling is to grow BHs from massive seeds, usually with $M_{\text{BH}} \gtrsim 10^5 M_\odot$, which are on or above the observed $M_{\text{BH}}-M_*$ relation, and to rely on regulations by the stellar or AGN feedback to drive them down to, and to move long, the relation. This approach is commonly adopted in cosmological hydrodynamic simulations (e.g. [Schaye et al. 2015](#); [Bower et al. 2017](#); [Weinberger et al. 2018](#); [Davé et al. 2019](#); [Habouzit et al. 2021](#); [Li et al. 2024](#)). However, whether and how such seeds can be bred remains unclear. Alternatively, one can start with seeds that are much less massive, and boost their growth via various mechanisms, such as mergers with star clusters brought in by dynamical friction, as suggested by [Dekel et al. \(2025\)](#), or the capture of the low- j sub-clouds that are regenerated continuously by injecting disturbance into the turbulent gas, as suggested by [Paper-I](#) (see their §4.3.3 and fig. 7).

In this paper, we extend the model presented previously in the papers of this series by developing prescriptions to seed and grow BHs in dark matter halos within the current cosmological context. We show how these prescriptions can be incorporated coherently into the two-phase scenario of galaxy formation, and be implemented semi-analytically into the halo merger trees to make model predictions. Our main goals are (i) to derive when, where and how BH seeds are bred in the early Universe and what masses these seeds have when they are produced; (ii) to understand the early growth of the seeds that eventually makes them supermassive, as well as their late growth that shapes the scaling relations observed at lower redshift; (iii) to differentiate effects from seeding and growing on the observed BH populations at different redshifts and to understand how halo assembly and environment affect the products; and (iv) to predict observable quantities and relations for halos, galaxies, star clusters

and (SM)BHs at different redshift, both statistically for populations and individually for representative objects. The paper is organized as follows. We start with descriptions of dark matter halos and their assembly histories in §2. In §3, we focus on the seeding procedure and discuss the physical concepts and prescriptions of how BH seeds form along with Pop-III star clusters. In §4, we focus on the growing procedure, and present concepts and prescriptions of how BHs grow via different channels. Analytical estimates are provided as much as possible so that intuition and order-of-magnitude quantification can be obtained without diving into the details. In §5, we first synthesize the physical prescriptions semi-analytically by implementing them into halo merger trees. We then present results predicted by the model and make comparisons with observations. In §6, we summarize our main results, and discuss their implications and potential follow-up investigations prompted by them.

2 DARK MATTER HALOS AND THEIR ASSEMBLY

A key component in the current Λ CDM paradigm of galaxy formation is the assembly of dark matter halos – dynamically stable clumps produced by gravitational instability of the cosmic density field, within which galaxies form and evolve. As we will show later, the early assembly of dark matter halos at high redshift is crucial for determining where and when the formation of the first generation of stars and the seeds of black halos occurs (§3), as well as the bursty growth of these seeds that is essential for them to grow into SMBHs (§4). Here we briefly summarize the properties and assembly histories of dark matter halos that are relevant to the description of our model.

Throughout this paper, we adopt a flat Λ CDM cosmology, with cosmological quantities defined as follows. The critical density at $z = 0$ is written as $\rho_{c,0} = 3H_0^2/(8\pi G)$, where $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ is the Hubble constant at the present day, and we assume $h = 0.6774$. The critical density at $z > 0$ is given by $\rho_c(z) = \rho_{c,0}E^2(z)$, where $E(z) = [\Omega_{m,0}(1+z)^3 + \Omega_{\Lambda,0}]^{1/2}$, and we assume the density parameters in the non-relativistic matter component, $\Omega_{m,0} = 0.3089$, and in the dark energy component (assumed to be a cosmological constant), $\Omega_{\Lambda,0} = 0.6911$. The cosmic baryon fraction: $f_b \equiv \Omega_{b,0}/\Omega_{m,0} \approx 0.157$, while the cosmic baryon density at $z \geq 0$ is

$$\rho_b(z) = \rho_{c,0}\Omega_{b,0}(1+z)^3 = \left[4.19 \times 10^{-28} \text{ g cm}^{-3}\right] (1+z)_{10}^3, \quad (1)$$

with $(1+z)_{10} \equiv (1+z)/10$. Grey segments in Fig. 6a shows the cosmic mean baryon density at three different redshifts. It varies from about 10^{-4} cm^{-3} at $z = 9$ to 10^{-7} cm^{-3} at $z = 0$, reflecting the factor $(1+z)^3$ that encodes the expansion of the Universe. The Gaussian initial density field has a power spectrum $P(k) \propto k^n$ with $n = 0.9667$ and an amplitude specified by $\sigma_8 = 0.8159$. The values of cosmological parameters quoted above are based on those given by Planck Collaboration et al. (2016).

A dark halo is defined by a spherical volume of virial radius R_v within which the mean mass density is $\Delta_v = 200$ times the critical density of the Universe (e.g. Navarro et al. 1997). The virial density, virial radius, virial velocity, and virial temperature are then given respectively by

$$\begin{aligned} \rho_v &\equiv \Delta_v \rho_{c,0} E^2(z) \\ &\approx \left[7.91 \times 10^{-3} \text{ M}_{\odot} \text{ pc}^{-3}\right] (1+z)_{10}^3; \end{aligned} \quad (2)$$

$$\begin{aligned} R_v &\equiv \left(\frac{2GM_v}{\Delta_v}\right)^{1/3} H^{-2/3}(z) \\ &\approx [4.57 \text{ kpc}] M_{v,9.5}^{1/3} (1+z)_{10}^{-1}; \end{aligned} \quad (3)$$

$$\begin{aligned} V_v &\equiv (\Delta_v/2)^{1/6} [GM_v H(z)]^{1/3} \\ &\approx \left[54.5 \text{ km s}^{-1}\right] M_{v,9.5}^{1/3} (1+z)_{10}^{1/2}; \end{aligned} \quad (4)$$

$$\begin{aligned} T_v &= 0.75 \times \frac{\mu m_p}{2k_B} V_v^2 \\ &\approx \left[1.62 \times 10^5 \text{ K}\right] M_{v,9.5}^{2/3} (1+z)_{10}. \end{aligned} \quad (5)$$

where the first line of each expression is the definition of the quantity; the approximation in the second line is valid at $z \gg 1$; $M_{v,9.5} \equiv M_v/(10^{9.5} \text{ M}_{\odot})$. The expression of virial temperature is given by e.g. Bryan & Norman (1998) and Wise & Abel (2007), with an additional factor 0.75 calibrated by Fernandez et al. (2014) using their hydro simulations, m_p is the proton mass, and we assume the mean molecular weight $\mu = 1.2$. We model the density profiles of halos using the NFW form:

$$\rho(r) \propto \frac{1}{x(1+x)^2}, \quad (6)$$

where $x \equiv r/R_s$, with R_s being a scale radius related to R_v by the concentration parameter, $c_v = R_v/R_s$ (Navarro et al. 1997). As shown by Zhao et al. (2009), the value of c_v of a halo is closely related to the mass assembly history of the halo. Finally, we define the dynamical timescale of a halo as

$$t_v \equiv \frac{R_v}{V_v} = \frac{1}{10H(z)} \approx 81.9 \text{ Myr} (1+z)_{10}^{-3/2}. \quad (7)$$

As one can see, at a given redshift, ρ_v and t_v depend only on cosmology, while R_v , V_v and T_v depend on both cosmology and M_v . Thus, if we know the assembly history of a halo, i.e. its M_v as a function of z , we can follow its properties over its history. As shown by calibrations using numerical simulations (e.g. Dekel et al. 2013), the average accretion rate of a virialized halo of mass M_v at $z \gg 1$ can be approximated as

$$\dot{M}_v \approx \left[10.3 \text{ M}_{\odot} \text{ yr}^{-1}\right] M_{v,9.5}^{1.14} (1+z)_{10}^{2.5}. \quad (8)$$

This can be integrated to obtain the average assembly histories and provides a guidance for the growth of individual halos. For reference, Fig. 1 shows the median assembly histories of halos of given masses at $z = 0$, in terms of M_v (panel a), c_v (panel b), R_v (panel c), V_v and T_v (panel d), and \dot{M}_v (panel e), as functions of z (or ρ_v). These curves are obtained by generating a large sample of halo merger trees with any given root mass at $z = 0$ by a Monte Carlo algorithm based on the extended Press-Schechter (EPS) formalism (see Appendix A1 for details), extracting the main branches from the trees, and computing the median of each property at each redshift. These curves show how the properties of a halo evolve along its assembly history to the present day. In particular, for halos with present-day mass $M_{v,z=0} > 10^{12} \text{ M}_{\odot}$, their progenitors at $z \gtrsim 15$ are quite similar, with mass $M_v \sim 10^8 \text{ M}_{\odot}$, $R_v \sim 1 \text{ kpc}$, $V_v \sim 20 \text{ km s}^{-1}$ and $T_v \sim 10^{4.5} \text{ K}$. These ‘mini-halos’, with a typical density $\rho_v \sim 0.1 \text{ M}_{\odot} \text{ pc}^{-3}$, are places to breed black-hole seeds, as we will see below. Panel b shows the halo concentration c_v of the NFW profile (Navarro et al. 1997) as a function of z . As suggested by Zhao et al. (2003, 2009), the assembly history of a halo can be separated into two regimes: an earlier fast-accretion regime, where V_v increases rapidly with time

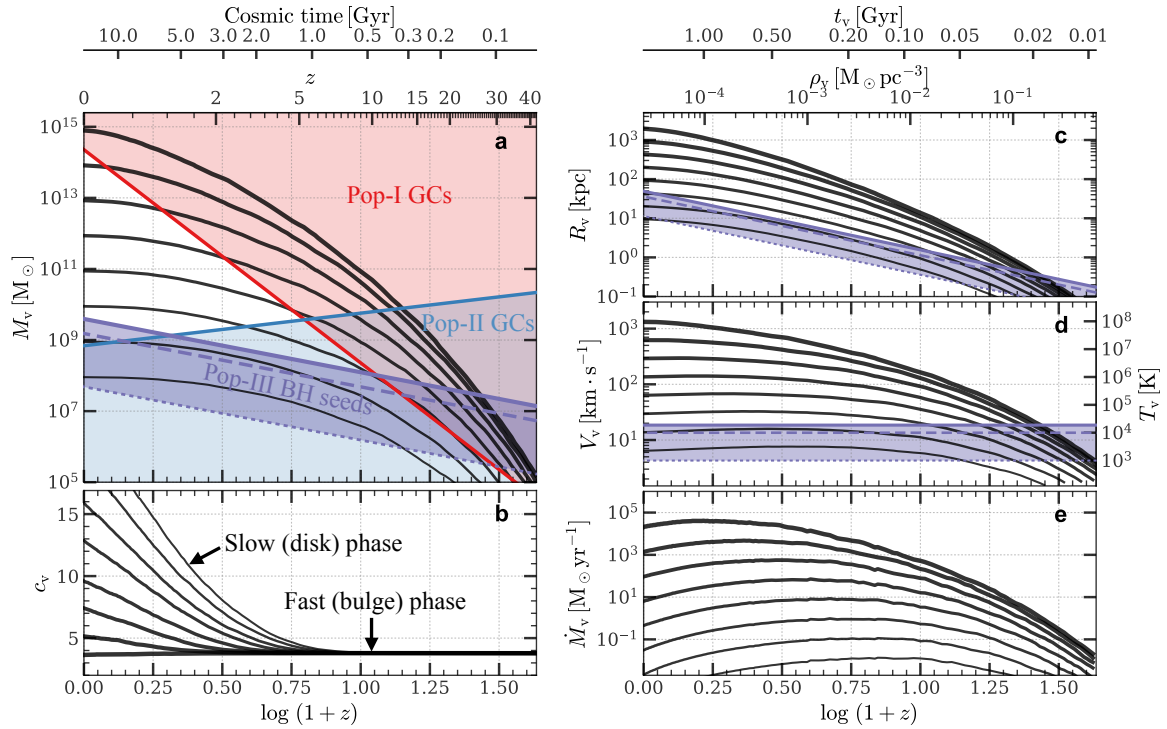


Figure 1. Assembly histories of dark matter halos. Here we show the evolution of different virial properties: **a**, mass (M_v); **b**, concentration (c_v); **c**, radius (R_v , in physical scale); **d**, velocity and temperature (V_v and T_v); **e**, accretion rate (\dot{M}_v). The assembly histories are computed along the main branches of halos at $z = 0$ using the EPS-based merger trees with Monte Carlo sampling (see Appendix A1), and averaged over large sets of halos with given $M_{v,z=0}$ from $10^8 M_\odot$ to $10^{15} M_\odot$ (**black curves** from thin to thick). **Purple bands** in **a**, **c** and **d** indicate the regime where the first collapse of an SGC is expected to be triggered so that Pop-III stars can form and BH seed can be bred, with the associated purple **dotted**, **dashed** and **solid** lines representing M_{v,H_2} (the mass for H_2 -cooling to be effective; Eq. 14), $M_{v,H}$ (the mass for H-cooling to be effective; Eq. 16), and $f_b^{-1} M_{J,\max}$ (the mass for the SGC with strong turbulence support to be Jeans unstable; Eq. 20), respectively. Depending on halo mass, assembly rate and UV background, BH seeds are expected to be bred at different locations within the purple bands (see §3.1 for details). For reference, **red** and **blue** regions in **a** show the regimes where Pop-I and Pop-II GCs can actively form, respectively, according to Paper-III. This figure summarizes the properties of halos over wide ranges of mass and redshift, which will be used as input of our model to determine when, where and how the first generation of stars and BH seeds form, as well as their post-seeding growth. See §2 for details.

and the structure of halo grows nearly self-similar with a constant concentration parameter $c_v = c_f \approx 4$; a later slow-accretion regime, where V_v changes slowly with time and c_v becomes larger than c_f as the halo grows its outskirts. Thus, a halo during the slow phase has a stabilized inner core formed during the fast phase and growing outskirts, separated by a radius of $R_f \approx c_f R_s$. This two-phase assembly of dark matter halos has important implications for galaxy formation in dark matter halos, as detailed in Paper-I, as well as for the seeding and growing of SMBHs, as we will see below.

In addition to the assembly of dark matter halos, environments of halos can also affect the formation of BH seeds through radiation exposure (see §3.1.3) and metal contamination (see §3.2). Environmental variations are not well captured by the assembly of individual halos, as halo assembly and structure properties correlate only weakly with the environment (see fig. 9 of Chen et al. 2020; Gao & White 2007; Wang et al. 2011). To model environmental variations more accurately, we use halos from a cosmological N-body simulation, TNG100-1-Dark, which is conducted as a part of the IllustrisTNG project (Pillepich et al. 2018; Nelson et al. 2019). This simulation uses a periodic box with a side length of $L_{\text{box}} = 75 h^{-1} \text{Mpc}$, 1820^3 dark matter particles, each with a mass of $6.0 \times 10^6 h^{-1} M_\odot$, and a Plummer-equivalent gravitational softening length that varies from $1 h^{-1} \text{kpc}$ at high z to $0.5 h^{-1} \text{kpc}$ at low z . Halos and subhalos are identified using the friends-of-friends (FoF) algorithm with a scaled linking length of 0.2 (Davis et al. 1985) and the SUBFIND algorithm

(Springel et al. 2001; Dolag et al. 2009), respectively. The lower limit for the FoF halo mass is about $2 \times 10^8 h^{-1} M_\odot$. Subhalo merger trees are constructed using the SUBLINK algorithm (Springel 2005; Boylan-Kolchin et al. 2009; Rodriguez-Gomez et al. 2015). The main progenitor of a subhalo is defined as the one with the most massive history among all progenitors (De Lucia & Blaizot 2007; Rodriguez-Gomez et al. 2015). The central subhalo of an FoF halo is defined as the one with the most massive history among all subhalos within the halo, and the main branch of an FoF halo corresponds to the main branch of its central subhalo. The specific run, TNG100-1-Dark, used here ensures a balance between statistical robustness and resolution in halo assembly histories.

The mass resolution of the cosmological N-body simulation adopted here is insufficient to resolve mini-halos needed to model the first generation of stars and BH seeds. To overcome this problem, we use the EPS-based Monte Carlo merger trees to extend the simulated merger trees down to the mini-halo regime. Appendix A1 details the extension algorithm and shows examples of simulated halo assembly histories supplemented with extended trees.

Similar to the models of Lu et al. (2014), the sample of subhalo merger trees used in this paper are those rooted in subhalos with host halo mass $M_v \geq 10^{9.5} M_\odot$, evaluated for central subhalos at $z = 0$ or satellite subhalos at their infall times. Subhalo above this mass limit are expected to grow above the atomic-cooling limit at some epochs (see below for the discussion of Fig. 1), and the choice of this limit

Cosmological context of black-hole seeding

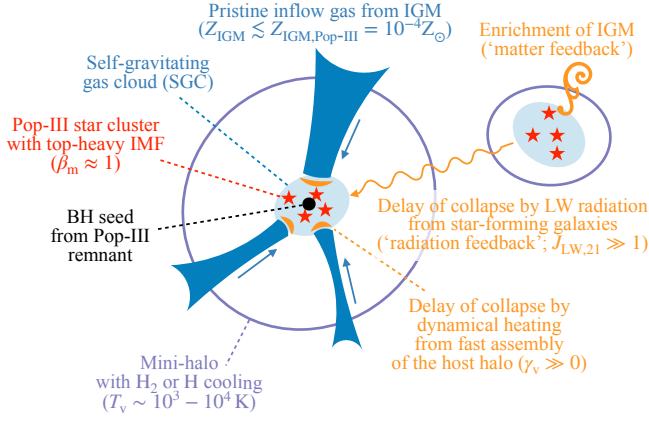


Figure 2. Schematic diagram summarizing the cosmological context of BH seeding. Pristine gas in a mini-halo that has reached the H_2 -cooling limit ($T_v \approx 10^3$ K; Eq. 14) can start to collapse to form a self-gravitating cloud (SGC) at the halo center (§3.1.1), contract further to form a Jeans core, and fragment to form a Pop-III star cluster. A light-weight BH seed can be bred when the most massive star (dominant star) in the cluster ends its life by core-collapse supernova (CCSN; §3.1.2). Gas collapse can be delayed to about the atomic-cooling limit ($T_v \approx 10^4$ K; Eq. 16) by ‘heating’ sources such as Lyman-Werner (LW) radiation from star-forming galaxies (§3.1.3) and/or dynamical heating from fast assembly of the host halo (§3.1.4), which also raise the mass of dominant star so that pair-instability supernova (PISN) or direct-collapse black hole (DCBH) can appear and leave a more massive seed (§3.1.5). A radiation-regulated scenario is proposed here to model the initial mass function (IMF) of Pop-III stars, and the mass spectrum of BH seeds is determined accordingly by following the stellar evolution of these stars (§3.2). The generation of Pop-III stars is prevented once the gas environment is enriched to the Pop-II level by the matter feedback from galaxy formation. Our scenario suggests that the environmental effects – concepts that have been built for galaxy formation at low z , have been present ubiquitously in the early Universe at $z \gtrsim 20$. See §3 for details.

thus ensures that galaxies with substantial star formation after the cosmic reionization are safely included in our analysis. Halos that do not grow above this limit may also have some star formation when, for example, the intergalactic medium (IGM) has high neutral fraction so that hard UV radiation is shielded (e.g. [Dijkstra et al. 2004](#); [Benson et al. 2006](#); [Ma et al. 2025b](#)), and thus may leave some extremely low-mass fossil galaxies at low z that are missed by our model.

The formation of virialized halos causes a significant jump of the density with respect to the density at large scale (see §4.1 for a thorough discussion of the cosmic hierarchy), setting the boundary condition for the subsequent gas evolution, star formation and black hole seeding/accretion, as we will discuss in the following.

3 BREEDING BLACK-HOLE SEEDS

As a dark matter halo forms, gas within the halo is heated by the virialization process, and must first cool down so as to collapse and form stars and BHs. The competition between heating and cooling processes thus determines when, where and how the first generation of stars and BHs forms. The seeding procedure follows the assembly of individual dark matter halos described in the last section, identifies the key processes of cooling and heating, and determines the key conditions for the gas to cool and collapse. The first generation of

stars in a halo is thus expected to form once these key conditions are all met. As we will show, the properties of the stellar population, including the masses of the dominant stars, is regulated by feedback processes. The fate of the first generation of stars is then determined according to the evolution of the stellar population, from which the masses and other properties of the BH seeds can be obtained. The properties of the stellar population and BH seeds thus encode the imprints of halo mass, assembly history and environment, as detailed in the following. For convenience, we illustrate the key components of the seeding model in Fig. 2.

3.1 Producing black-hole seeds in mini-halos

3.1.1 Formation of self-gravitating clouds

If the angular-momentum support is negligible, as is the case when the fraction of the cooled gas in a halo is not much lower than the universal baryon fraction, the collapse of gas towards the halo center is expected to lead to the formation of an SGC supported by supersonic turbulence, as shown in [Paper-I](#). For a halo with a gas fraction $f_g \equiv M_{\text{SGC}}/M_v$, the gas becomes self-gravitating after contracting by a factor $\sim 1/f_g$, and the SGC eventually becomes virialized at a radius of $f_r R_v$, where f_r is a coefficient to be determined. According to [Paper-III](#) (see their §2.2), the radius, density, surface density and circular velocity of such an SGC can be written as

$$R_{\text{SGC}} = f_r R_v \approx [183 \text{ pc}] f_{r,0.04} M_{v,9.5}^{1/3} (1+z)_{10}^{-1}; \quad (9)$$

$$n_{\text{SGC}} = \frac{f_g}{\mu m_p f_r^3} \rho_v \approx [656 \text{ cm}^{-3}] f_{g,0.157} f_{r,0.04}^{-3} (1+z)_{10}^3; \quad (10)$$

$$\begin{aligned} \Sigma_{\text{SGC}} &\sim \frac{M_{\text{SGC}}}{R_{\text{SGC}}^2} \\ &\approx [1.49 \times 10^4 \text{ M}_\odot \text{ pc}^{-2}] f_{g,0.157} f_{r,0.04}^{-2} M_{v,9.5}^{1/3} (1+z)_{10}^2; \end{aligned} \quad (11)$$

$$\begin{aligned} V_{\text{SGC}} &= \sqrt{\frac{GM_{\text{SGC}}}{R_{\text{SGC}}}} \\ &\approx [108 \text{ km s}^{-1}] f_{g,0.157}^{1/2} f_{r,0.04}^{-1/2} M_{v,9.5}^{1/3} (1+z)_{10}^{1/2}, \end{aligned} \quad (12)$$

respectively, and the free-fall timescale ($t_{\text{ff,SGC}}$) and dynamical timescale ($t_{\text{dyn,SGC}}$) are

$$\begin{aligned} t_{\text{ff,SGC}} &= \frac{\pi}{\sqrt{8}} t_{\text{dyn,SGC}} = \sqrt{\frac{3\pi}{32G\rho_{\text{SGC}}}} = [1.48 \text{ Myr}] n_{\text{SGC},3}^{-1/2} \\ &= [1.84 \text{ Myr}] f_{g,0.157}^{-1/2} f_{r,0.04}^{3/2} (1+z)_{10}^{-3/2}, \end{aligned} \quad (13)$$

where the SGC density $\rho_{\text{SGC}} = \mu m_p n_{\text{SGC}}$, and we write $n_{\text{SGC}} = 10^3 \text{ cm}^{-3} n_{\text{SGC},3}$, $f_g = 0.157 \times f_{g,0.157}$ and $f_r = 0.04 \times f_{r,0.04}$.

The collapse of gas within a dark matter halo and the formation of the SGC mark a jump of gas density relative to that of the halo gas (see §4.1 for a thorough discussion of the cosmic hierarchy). Depending on its internal properties and environment, an SGC can fragment to form sub-clouds which, in turn can collapse further to form compact objects of different masses, such as dense gas cores and stars. Particularly interesting to us are the formation of massive dense cores and the first massive stars at high z , and the possibility for them to produce BH seeds. In the rest of this subsection, we describe in more detail the formation of SGCs in low-mass halos at high z and discuss potential channels of BH seed production in SGCs.

3.1.2 *Gas collapse in mini-halos at high z*

At high z where the gas was pristine, the collapse of the halo gas can occur in low-mass mini-halos where H_2 cooling can keep the gas temperature at a level of $T \approx T_{\text{H}_2} = 10^3 \text{ K}$ (Latif & Khochfar 2019). Thus, as the virial temperature of a mini-halo reaches $T_v = T_{\text{H}_2}$, the gas starts to contract due to H_2 cooling. A more quantitative analysis by Yoshida et al. (2003) and Glover (2013), based on the production of H_2 and the required fraction of H_2 for halo gas to cool within a dynamical (free-fall) timescale ($t_{\text{cool}} \sim t_{\text{dyn}} \sim t_{\text{ff}}$) gives a similar value for the critical temperature. Using T_v given by Eq. (5), we get the critical halo mass for the gas to contract:

$$M_{v,\text{H}_2} = \left[3.87 \times 10^5 M_\odot \right] (1+z)_{25}^{-3/2} T_{\text{H}_2,3}^{3/2}, \quad (14)$$

where $(1+z)_{25} \equiv (1+z)/25$, $T_{\text{H}_2,3} \equiv T_{\text{H}_2}/(10^3 \text{ K})$. The redshift $(1+z) = 25$ gives a halo dynamical timescale $\approx 20.7 \text{ Myr}$, which is roughly the time for the cooled gas to become self-gravitating. According to Eqs. (9)–(13), the SGC produced has mass $M_{\text{SGC}} = f_b M_{v,\text{H}_2} = 6.08 \times 10^4 M_\odot$, radius $R_{\text{SGC}} = 3.63 \text{ pc}$, density $n_{\text{SGC}} = 1.03 \times 10^4 \text{ cm}^{-3}$, circular velocity $V_{\text{SGC}} = 8.49 \text{ km s}^{-1}$, and free-fall timescale $t_{\text{ff,SGC}} = 0.463 \text{ Myr}$. Purple dotted lines in Fig. 1 show M_{v,H_2} as a function of redshift (panel a) assuming $T_{\text{H}_2,3} = 1$, and the corresponding R_v (panel c), V_v and T_v (panel d) obtained from M_{v,H_2} using Eqs. (3), (4) and (5).

The temperature, $T = T_{\text{H}_2}$, corresponds to a thermal sound speed of $c_s \approx 3.4 \text{ km s}^{-1}$. The implied Jeans mass of the SGC,

$$M_{\text{J,H}_2} = \frac{\pi^{5/2}}{6} \frac{c_s^3}{\sqrt{G^3 \rho_{\text{SGC}}}} \\ = 2.30 \times 10^4 M_\odot f_{\text{g},0.157}^{-1/2} f_{\text{r},0.04}^{3/2} (1+z)_{25}^{-3/2} T_{\text{H}_2,3}^{3/2}, \quad (15)$$

sets the lower mass limit for the gas to collapse and fragment. The fact that this Jeans mass is comparable to the mass of the SGC and that the cooling by H_2 is not rapid ($t_{\text{cool}} \sim t_{\text{dyn}}$) implies that the contraction of the SGC proceeds globally without leading to fragmentation until other cooling channels are open. For gas with a primordial composition, where cooling from metals and dust grains is insignificant, the whole SGC cools and contracts gradually to a high density before fragmentation can occur.

When the density in the core of an SGC reaches $n = n_{\text{core}} \approx 10^6 \text{ cm}^{-3}$, three-body reactions rapidly cools the gas, boosting the production of H_2 and making the core unstable (see §3.3 of Latif & Khochfar 2019). The Jeans mass in the core is $M_{\text{J,core}} \approx 1.88 \times 10^3 M_\odot \ll M_{\text{g,cool}}$, corresponding to a size of $R_{\text{core}} \approx 0.25 \text{ pc}$ for the density in question. Within such a Jeans core, conditions for star formation are satisfied, and fragmentation starts from the center of the core, leading to the formation of a central (proto-)star followed by the formation of other stars. Subsequent rapid inflow and frequent mergers increase the mass of the central star, which may lead to a top-heavy IMF for the stellar population in the core, with the total stellar mass dominated by the central star (see e.g. Stacy et al. 2016; Latif et al. 2022b). As suggested by numerical simulations, the central star, with a typical mass of a few $10 M_\odot$, dies within a timescale of Myr due to core collapse supernova (CCSN) explosion and leaves behind a BH with mass slightly lower than the stellar mass (see e.g. §3.1 of Costa et al. 2023). A ‘light-weight’ seed to grow SMBH is then produced.

The above discussion only takes into account radiative cooling, ignoring potential heating sources that may be present. In the presence of heating, the contraction of the halo gas to form an SGC, and the fragmentation and collapse of the SGC to form stars, are delayed so that the mass of the Jeans core is increased. The central part of the

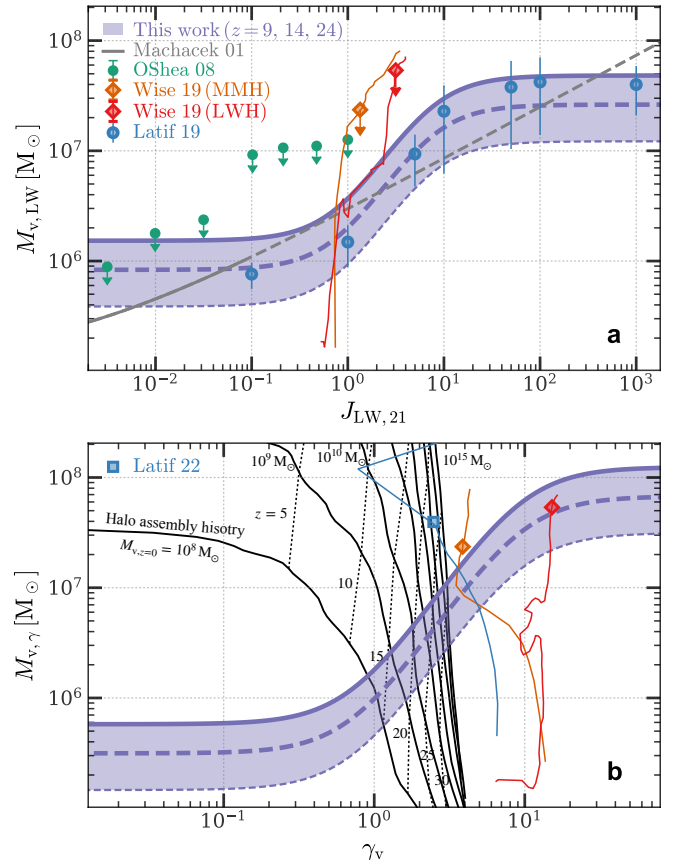


Figure 3. Delay of gas collapse in mini-halos. Here we show the threshold mass of halo expected for the gas to collapse, under **a**, different strength of UV background ($J_{LW,21}$, measured in the LW band) and **b**, different specific rate of halo accretion (γ_v). The former measures the reduction of gas cooling due to the reduction of H_2 production, while the latter measures the dynamical heating due to fast accretion. In both panels, solid and dotted purple curves show our parameterization of the threshold mass at $z = 9, 14$ and 24 , respectively. Black curves in **b** show the assembly histories of halos in the M_v - γ_v plane, solid for halos with given masses at $z = 0$ (from 10^9 to $10^{15} M_\odot$, each separated by 1 dex), and dashed for contours at given z (from 5 to 30), obtained by averaging a large set of EPS-based Monte Carlo merger trees (see Appendix A1). The SGC in a pristine halo with dynamical heating thus collapses and forms its first generation of stars once the trajectory of the halo passes the purple band. Colored markers show simulation results obtained by O’Shea & Norman (2008), Wise et al. (2019), Latif & Khochfar (2019) and Latif et al. (2022a). The halos used by O’Shea & Norman (2008) and Wise et al. (2019) are extremely massive ones, dominated by the effect of dynamical heating, and thus we show their results as upper limits in **a**. Colored curves associated with the markers for Wise et al. (2019) and Latif et al. (2022a) are the histories of LW background (**a**) and halo assembly (**b**). Grey curve in **a** shows the fitting result, $M_{v,\text{cool}}/M_\odot = 1.25 \times 10^5 + 2.86 \times 10^6 J_{LW,21}^{0.47}$, obtain by Machacek et al. (2001) using their simulations with $J_{LW,21} \lesssim 0.1$, with dashed segment indicating the extrapolation. See §3.1 for details.

Jean core may then collapse directly to form a BH with mass greater than a few $100 M_\odot$, producing a ‘heavy-weight’ seed to grow SMBH. Two sources, one from UV radiation in the environment and the other from the energy of the turbulence associated with the fast-accretion regime of the host halo, are particularly relevant to our problem and their effects are described below.

3.1.3 Delay of collapse by UV radiation

For a given halo, UV radiation produced by young stars in nearby galaxies can provide external feedback that suppresses gas cooling and collapse in the halo. At high z when the Universe is still largely neutral, neutral hydrogen in the IGM can effectively absorb UV photons above 13.6 eV, the Lyman limit of H, but the IGM remains optically thin for photons below that limit. This leads to significant attenuation only in resonance frequencies, producing a saw-toothed spectrum (see e.g. §2.1 of Haiman et al. 2000). Such photons can dissociate H_2 via the Solomon process, $H_2 + \gamma \rightarrow H_2^* \rightarrow H + H$, where H_2^* is any of the resonance states above 11.2 eV (see e.g. §2 of Machacek et al. 2001). As H_2 is needed to cool the gas, UV photons within the window of 11.2–13.6 eV, usually referred to as the Lyman-Werner (LW) radiation, can suppress gas cooling and the formation of stars and BHs in the halo.

The intensity of the LW radiation, J_{LW} , is defined as the energy in the LW band received at a location per unit time per unit area per unit frequency and per unit solid angle. As a convention, we define $J_{LW} = J_{LW,21} 10^{-21} \text{erg s}^{-1} \text{cm}^{-2} \text{Hz}^{-1} \text{sr}^{-1}$. The threshold mass for a H_2 -cooling mini-halo to collapse is expected to increase with increasing J_{LW} . If the radiation is strong enough ($J_{LW,21} \gg 1$), the collapse of the halo gas to form an SGC can be delayed until atomic cooling kicks in, namely when $T_V \approx T_H \approx 10^4$ K. The halo mass corresponding to this temperature can be derived in a way similar to Eq. (14):

$$M_{V,H} = \left[1.22 \times 10^7 M_\odot \right] (1+z)^{-3/2} T_{H,4}^{3/2}, \quad (16)$$

where $T_{H,4} \equiv T_H/(10^4 \text{K})$. According to Eqs. (9)–(13), the SGC produced has mass $M_{SGC} = f_j M_{V,H} = 1.92 \times 10^6 M_\odot$, radius $R_{SGC} = 11.5 \text{pc}$, circular velocity $V_{SGC} = 26.9 \text{km s}^{-1}$, and density and free-fall timescale the same as those of the H_2 -cooling SGC at a given redshift (see the texts below Eq. 14). Purple dashed lines in Fig. 1a, c and d show $M_{V,H}$ and the corresponding R_V and V_V (T_V), respectively, as a function of redshift. They are parallel to but higher than the corresponding H_2 -cooling lines. More generally, we can parameterize the halo mass for effective cooling to drive the formation of SGC as a function of J_{LW} as

$$M_{V,LW} = \frac{M_{V,H_2} + f_j M_{V,H}}{1 + f_j}, \quad (17)$$

where f_j depends on the intensity of the LW radiation irradiated on the halo,

$$f_j = \left(\frac{J_{LW,21}}{J_{\text{crit},21}} \right)^{\beta_j}. \quad (18)$$

We set $\beta_j = 1.5$ and $J_{\text{crit},21} = 7.5$ by default to match the results from zoom-in hydrodynamic simulations (see below). So modeled, $M_{V,LW}$ smoothly transits from the H_2 -cooling limit to the H-cooling limit at any given redshift. Fig. 3a shows our default results of $M_{V,LW}$ as a function of J_{LW} at different redshifts, in comparison to results obtained by other studies. Machacek et al. (2001) performed zoom-in hydrodynamic simulations for halos embedded in different levels of UV background with $J_{LW,21} \lesssim 0.1$. They found that the threshold halo mass ($M_{V,LW}$) required for the gas to collapse increases with J_{LW} , and fitted the $M_{V,LW}$ - J_{LW} relation with a power-law function (grey curve in Fig. 3a, with dashed segment indicating the extrapolation). Latif & Khochfar (2019) performed zoom-in simulations for 6 halos, selected at $z \approx 23$ –26 with $M_V = \text{a few times } 10^5$ – $10^6 M_\odot$, imposing different levels of UV radiation with $J_{LW,21} = 0.1$ –1000. The collapse of the gas is significantly delayed at high $J_{LW,21}$. The identified collapsing threshold $M_{V,LW}$ as a function of $J_{LW,21}$ is shown by blue markers in Fig. 3a, with error bars indicating the 1- σ dispersion among halos.

The threshold rises sharply at $J_{LW,21} \gtrsim 1$, consistent with the results of Machacek et al. (2001), and saturates to a constant values at $J_{LW,21} \gtrsim 10$. Given the wide range of J_{LW} covered by this set of simulations, we calibrate our parameterization in Eq. (18) with it, yielding the default parameters of β_j and $J_{\text{crit},21}$. For comparison, green markers show the results obtained by O’Shea & Norman (2008) using a set of zoom-in simulations for one halo with different levels of J_{LW} ; orange and red markers show the results obtained by Wise et al. (2019) using two zoom-in simulations for two halos, respectively, with two curves passing the markers showing the histories of LW radiation produced mainly by nearby star-forming galaxies impacting the halos. Their halos are selected to be extremely massive ones in cosmological volumes, thus experiencing fast accretion and expected to have strong dynamical heating (see §3.1.4). Thus, we show their results for the collapsing threshold as upper limits in Fig. 3a.

Our modeling of BH-seed breeding takes into account effects of UV radiation in the LW band, as described above. Other effects of the UV radiation, such as the detachment of H^- by photons of $> 0.76 \text{eV}$, can also affect the formation of H_2 . Unfortunately, a reliable model of such effects requires detailed modeling of the stellar spectrum energy distribution (SED), which depends on the star formation and chemical enrichment history of the emitting galaxy, and radiative transfer, which depends on the properties of the interstellar medium (ISM), circumgalactic medium (CGM) and IGM, both beyond the scope of this paper. A less sophisticated model of the UV radiation, as adopted in the hydrodynamic simulations of Latif & Khochfar (2019), is to assume that the emitted radiation has a black-body SED with some characteristic temperature in the range of 10^4 – 10^5 K. Our calibration using the results of Latif & Khochfar (2019) thus partially captures the H^- photodetachment. Agarwal et al. (2019) presented a more detailed model of the UV radiation by post-processing hydrodynamic simulations, synthesizing stellar SEDs, and adopting a critical curve for the DCBH in the 2-D plane of H_2 photodissociation and H^- photodetachment. Their results suggest that the net effect of the UV radiation on the delay of the SGC collapse is tightly correlated with the star formation rate (SFR) of emitting galaxies. This motivates us to use the SFR (or young stars) as a proxy of J_{LW} in the threshold variable f_j (Eq. 18) as an effective way to account for the net effect of the UV radiation. Since our model can predict star formation histories of individual galaxies in a cosmological volume, the information needed for the modeling of UV radiation is self-consistently included in the model. In Appendix A5, we will present the detailed steps for such a model. The UV radiation produced by young stars, and its effects on the delay of SGC formation, are the earliest ‘radiation feedback’ of galaxy formation to its environment.

3.1.4 Delay of collapse by turbulence associated with fast accretion

As discussed in Paper-I, the SGC in a fast-accreting halo is expected to be a turbulent system consisting of high-density, fast moving sub-clouds that support the structure of the SGC. The situation is the same here, because halos to form the first generation of compact objects are almost certainly in the fast-accreting regime, and because the low metallicity of the halo gas makes it hard for radiative cooling to fully dissipate the turbulence. Depending on the details of the gas collapse and the turbulence cascading, the partition between the kinetic and thermal energy can be different for different halos. Wise et al. (2019) reported results of two simulated halos at $z \approx 15$ as potential sites to form supermassive stars (SMSs). In each of the two cases, the external LW radiation is moderate, while the fast halo accretion heats the SGC and leads to the formation of a hot, thermally supported core with mass $\gtrsim 10^3 M_\odot$ (see their Extended

Data fig. 3). [Latif et al. \(2022a\)](#) reported results of one primordial halo at $z \approx 25$ as potential site of SMS formation. The halo is not exposed to any external LW radiation, but its fast accretion leads to supersonic turbulence that supports the entire SGC. The energy of the turbulence in these examples all comes from gravity. We thus refer to this channel of energy injection by the fast accretion of halos collectively as ‘dynamical heating’, regardless whether the injected energy is stored in kinetic or thermal form.

To model the effect of the dynamical heating, we first consider an extreme case where energy injection is so efficient that radiative cooling is inefficient to dissipate it. If the cooling timescale t_{cool} becomes comparable or longer than V_v/\dot{V}_v , the time scale for the potential energy of a halo to be totally flushed by infalling material, the gravitational energy associated with the mass accretion of the halo can generate a significant impact on the SGC over the cooling timescale. Thus, gas in a halo can collapse rapidly only when $t_{\text{cool}} \lesssim V_v/\dot{V}_v$. Equating the two timescales, we obtain a critical condition for the turbulence to affect the state of the SGC: $V_v/\dot{V}_v \approx t_{\text{cool}}$. Noticing that at the collapse of SGC, $t_{\text{cool}} \sim t_v$, and using Eq. (7) for t_v , we have

$$\gamma_v \equiv \frac{\dot{V}_v}{V_v H(z)} = -\frac{d \ln V_v}{d \ln(1+z)} \approx 10. \quad (19)$$

To incorporate the additional support to SGC by turbulence, we replace the sound speed in the Jeans mass (Eq. 15) with an effective velocity dispersion, $\sigma_{\text{eff}} = (c_s^2 + v_t^2)^{1/2}$. The exact value of the turbulent velocity dispersion v_t is unknown. As a conservative estimate, we take the value, $v_t \approx 20 \text{ km s}^{-1}$, obtained by [Latif et al. \(2022a\)](#) from their simulation of a fast-accreting halo with H_2 cooling unable to dissipate the turbulence. This value of v_t , together with c_s given by $T = T_{\text{H}_2}$ and the SGC density given by Eq. (10), leads to the following Jeans mass:

$$M_{\text{J,max}} = \left[4.94 \times 10^6 M_{\odot} \right] f_{g,0.157}^{-1/2} f_{r,0.04}^{3/2} (1+z)_{25}^{-3/2}. \quad (20)$$

The SGC thus starts to collapse rapidly only when it reaches this mass scale. The corresponding radius and circular velocity are $R_{\text{SGC}} = 15.7 \text{ pc}$ and $V_{\text{SGC}} = 36.8 \text{ km s}^{-1}$, respectively. The corresponding host halo mass, $M_v = M_{\text{J,max}}/f_b = 3.14 \times 10^7 M_{\odot}$, is slightly larger than $M_{v,\text{H}}$ given by Eq. (16), indicating that dynamical heating can delay the formation of the first generation of compact objects by a timescale similar to that implied by a strong UV radiation field. Purple solid lines in Fig. 1a, c and d show $M_v = M_{\text{J,max}}/f_b$ and the corresponding R_v and V_v (T_v), respectively, as a function of redshift. They are parallel to but higher than the corresponding H_2 -cooling lines, and close to the corresponding H-cooling lines.

More generally, a halo can be somewhere between the two extremes, the one without turbulence ($v_t \ll c_s$) and the other with strong turbulence ($v_t \gg c_s$). The Jean mass is thus between those given by Eq. (15) and Eq. (20), with a transition at $\gamma_v \approx 10$. We parameterize the halo mass at which the thermal and turbulent motion is no longer able to support the SGC against gravitational collapse as a function of γ_v :

$$M_{v,\gamma} = f_b^{-1} \frac{M_{\text{J,H}_2} + f_{\gamma} M_{\text{J,max}}}{1 + f_{\gamma}}, \quad (21)$$

where

$$f_{\gamma} = \left(\frac{\gamma_v}{\gamma_{\text{crit}}} \right)^{\beta_{\gamma}}, \quad (22)$$

and we adopt $\gamma_{\text{crit}} = 10$. The exact value for β_{γ} is not known, and we adopt $\beta_{\gamma} = 2$ as a demonstration of the effect of dynamical heating. Future calibrations using detailed hydrodynamic simulations

are needed to provide a more reliable estimate of γ_{crit} and β_{γ} , and to account for the synergy of dynamical heating with other heating processes (e.g. the UV radiation discussed in §3.1.3).

Purple curves in Fig. 3b show our model of the collapsing threshold, $M_{v,\gamma}$, at three different redshifts, as a function of halo specific growth rate, γ_v . The threshold from $z = 24$ to 9 covers a narrow band of ≈ 0.3 dex due to the dependence of the Jeans mass on SGC density (thus redshift). The delay of SGC collapse by dynamical heating can lead up to ≈ 2 dex increase in halo mass when $\gamma_v \geq 10$. Since halos with higher mass accrete faster than those with lower mass at a given redshift, formation of BH seeds in the progenitors of massive halos at present day are most likely affected significantly by dynamical heating. To see this, we take the EPS-based halo assembly histories used to produce Fig. 1, compute the specific growth rate γ_v , and average over halos with given $M_{v,z=0}$. Solid black curves in Fig. 3b show the M_v - γ_v trajectories of halos with different $M_{v,z=0}$, while each dashed curve links a snapshot of these trajectories at a given redshift. Halos with $M_{v,z=0} \geq 10^{12} M_{\odot}$ have very similar M_v and γ_v at $z \geq 15$, consistent with the assembly histories shown in Fig. 1. γ_v exhibits slow reduction with redshift for all halos at high z . The significant reduction of γ_v occurs earlier for halos with lower mass, consistent with the findings in [Paper-I](#) (see their fig. B1) that z_f shows negative correlation with $M_{v,z=0}$.

The expected epoch of collapse, as well as host-halo properties at this epoch, of an SGC being dynamically heated by fast accretion (without taking into account other heating sources) can be identified by the intersection point of the M_v - γ_v trajectory with the purple band in Fig. 3b. Since the slopes of equal-time contours (dashed black) are higher than that of the purple band in the M_v - γ_v plane, SGCs in massive halos are expected to collapse earlier than those in less massive halos. Note that this is only a conclusion averaged over ensembles of halos, and halo-to-halo variations can be significant. Halos with $M_{v,z=0} = 10^{11} M_{\odot}$ (hosting dwarf galaxies with $M_* \approx 10^9 M_{\odot}$ at $z = 0$) reach the collapsing threshold at $z \approx 25$ when M_v is a few times $10^6 M_{\odot}$ while those with $M_{v,z=0} \geq 10^{13} M_{\odot}$ (hosting BCGs at $z = 0$) are expected to collapse at $z \approx 30$ with a comparable M_v . The values of M_v at collapse for these halos are thus ≈ 1 dex higher than that of H_2 -cooling halos (M_{v,H_2} , Eq. 14), implying that much more baryon gas is available for the formation of stars and BHs in the SGCs. Meanwhile, such a delay of collapse also gives the SGCs more chance to be irradiated by the UV photons from nearby star-forming galaxies, causing further delay, as we will discuss and model in §3.1.5.

For comparison, we show by colored curves in Fig. 3b the assembly histories of three pristine halos whose baryonic processes were resolved by hydrodynamic simulations. Two of them are the same halos as those in panel a, studied by [Wise et al. \(2019\)](#), and one by [Latif et al. \(2022a\)](#). All of them are progenitors of extremely massive halos, and were predicted to seed DCBHs and then evolve to quasars at $z \approx 6$. γ_v of these halos are extremely large, reflecting an accretion faster than that expected from progenitors of halos with $M_{v,z=0} \approx 10^{15} M_{\odot}$. Dynamical heating is thus expected to be very strong in these halos, causing significant delay of the SGC collapse. These are indeed seen in these simulations, as shown by the colored markers in Fig. 3b that denote the collapses they identified, close to the prediction of our model.

3.1.5 Consequences of the delayed collapse

As shown in Fig. 3, both UV radiation and fast accretion can delay the collapse of SGCs, the formation of the first generation of stars and the breeding of BH seeds. The effects are described by the two

halo mass thresholds, $M_{v,LW}$ and $M_{v,\gamma}$, respectively. In the presence of both effects, halo mass must reach the higher threshold of the two,

$$M_{v,th} = \max [M_{v,LW}, M_{v,\gamma}] , \quad (23)$$

in order for the SGC to overcome the heating from the dominant source, cool down and collapse. If either J_{LW} or γ_v is sufficiently large, the collapse is delayed to the regime of atomic cooling (Eq. 16). Consequently, the total available baryon mass, $M_{g,th} \approx f_b M_{v,th}$, and the halo virial velocity, V_v , are much larger than those for a H₂-cooling halo.

The purple band in Fig. 1a summarizes the regime in the z - M_v plane where the first collapse of an SGC in a halo can take place. The dotted, dashed and solid curves represent M_{v,H_2} (the mass for the H₂-cooling to be effective; Eq. 14), $M_{v,H}$ (the mass for the H-cooling to be effective; Eq. 16), and $f_B^{-1} M_{J,max}$ (the halo mass for the SGC to be Jeans-unstable at $T = 10^3$ K and $v_t = 20$ km s⁻¹; Eq. 20), respectively. These masses are plotted on top of the median assembly histories of halos of different masses at $z = 0$ (black curves). Pop-III stars are expected to form in a primordial halo when the halo mass, $M_v(z)$, crosses M_{v,H_2} (dotted line) and enters the Pop-III band limited by J_{LW} (dashed line) or γ_v (solid line). Note that the region for Pop-III stars in the z - M_v plane is almost fully (partly) covered by that marked Pop-II (Pop-I) GC, the region in which Pop-II (Pop-I) globular clusters are expected to form (see Paper-III). Thus, only halos that have not been self-enriched by previous generations of stars or externally enriched by star formation in nearby galaxies can host Pop-III stars, while the others are likely to host globular clusters. In the progenitors of halos with $M_{v,z=0} \approx 10^9$ – $10^{10} M_\odot$, Pop-III stars can form at $z \lesssim 10$ through the first collapse of the SGC provided that the IGM around these halos has not been enriched and the UV radiation and/or dynamical heating are sufficiently strong to delay the collapse to that low redshift. These Pop-III stars of late-formation may thus be observable in JWST surveys, such as those conducted recently by Wang et al. (2024b), Hsiao et al. (2025), Suess et al. (2024), Naidu et al. (2024) and Fujimoto et al. (2025). The enrichment history of both the ISM and IGM turns out to be critical in disentangling the two populations of stars in the z - M_v plane, as to be discussed in Appendix A6. For reference, purple bands in Fig. 1c and d show the first-collapse regime in the z - R_v and z - V_v (z - T_v) planes, respectively.

The free-fall timescale of an SGC at its first collapse is about 1 Myr (see the texts below Eq. 14), independent of halo mass (Eq. 13). This timescale is shorter than the ‘supernova-free’ timescale, $t_{snf} \approx 1$ Myr (Chen et al. 2025a; Dekel et al. 2023). Thus, supernova (SN) feedback is expected to be unimportant for the first collapse of the SGC. This is particularly true for the Jeans core that has $n \gtrsim 10^6$ cm⁻³, corresponding to an even shorter t_{ff} of $\lesssim 0.05$ Myr. Without other types of feedback, the upper limit of the gas accretion rate onto the central star can then be estimated as

$$\dot{M}_g \leq \frac{f_b M_{v,th}}{t_{ff,SGC}} \leq \frac{M_{J,max}}{t_{ff,SGC}} \approx 10.7 M_\odot \text{yr}^{-1} , \quad (24)$$

independent of redshift due to the cancellation of the redshift dependence in $M_{J,max}$ (Eq. 20) and $t_{ff,SGC}$ (Eq. 13). As suggested by numerical simulations of the formation of supermassive stars (e.g. Hosokawa et al. 2012, 2013), a high accretion rate at $\gtrsim 10^{-2} M_\odot \text{yr}^{-1}$ can lead to very different evolution in the stellar structure than a lower rate. After the triggering of the Kelvin-Helmholz (KH) contraction in the core, the radiation due to the release of gravitational energy causes the outer envelope to expand, leading to the formation of a stellar object with inhomogeneous structure and non-monotonic

density profile. Hydrogen burning can be ignited in the core, but provides only minor contribution to the total energy flux. At a stellar mass $M_* \sim 100 M_\odot$, the outer radius is related to M_* as

$$R_* \approx 2.6 \times 10^4 R_\odot \left(\frac{M_*}{100 M_\odot} \right)^{1/2} = 12 \text{ au} \left(\frac{M_*}{100 M_\odot} \right)^{1/2} , \quad (25)$$

independent of the accretion rate. The stellar object, with its effective temperature locked at a low value of $T_{\text{eff}} \approx 5000$ K due to strong temperature dependence of the H⁻ absorption, does not evolve to the main sequence. The UV luminosity at such a surface temperature is low, and the radiative feedback ineffective. The stellar object can then continue to accrete and grow significantly in mass, up to $\sim 10^3$ – $10^5 M_\odot$, until the density of the stellar envelop drops to a level that significantly reduces the opacity of H⁻. The upper limit of \dot{M}_g given by Eq. (24) for the central stellar object in an atomic-cooling halo is thus sufficient for it to grow into a supermassive star within a timescale less than 1 Myr, before an SN can detonate. Idealized simulations (Choi et al. 2013) and cosmological zoom-in simulations (Latif et al. 2013; Wise et al. 2019; Latif et al. 2021, 2022a) have shown that such high accretion rates are indeed achievable in a real physical setting and in the cosmological context. Such an SMS is expected to collapse when the gas pressure is no longer able to counterbalance its weight, leading to the formation of an intermediate-mass, direct-collapse black hole (DCBH), with a BH mass comparable to the SMS.

The above estimate ignores potential suppression of the accretion rate by feedback, angular momentum support and fragmentation of the gas to form low-mass stars. It thus only represents an upper limit of the mass that a DCBH can reach in a mini-halo. In the following, we consider a more realistic model that takes into account the effects of radiative feedback from the central star, and use it to derive the mass spectra of Pop-III stars and BH seeds expected to form during the first collapse of an SGC in a mini-halo.

3.2 Radiation-regulated mass function of Pop-III stars and the mass spectrum of black-hole seeds

The description presented above provides the avenue to model the mass of individual Pop-III stars expected to form in halos where the gas state is regulated by different cooling and heating processes. In this subsection, we first use the results to construct a model for the initial mass function (IMF) of Pop-III stars. We then combine the results with assumptions of stellar evolution to obtain the mass spectrum of BH seeds.

As seen above, the slow cooling of primordial gas has important implications to the star formation process. For the radiative cooling to effectively dissipate the energy generated by the gravitational collapse, the SGC has to contract to achieve a sufficiently high density. Gas fragmentation and star formation, which are possible only when the SGC can cool effectively, are thus delayed until the gas density is sufficiently high (see e.g. fig. 9 of Omukai et al. 2008). The situation is similar to that in the formation of Pop-II GCs in metal-poor ($Z \sim 10^{-2} Z_\odot$) sub-clouds with $n \gtrsim n_{snf} = 10^{3.5}$ cm⁻³ (Chen et al. 2025a), except the more extreme condition produced by the fact that the temperature does not drop significantly even when the density reaches $n \gtrsim n_{\text{core}} = 10^6$ cm⁻³ in a Pop-III gas (see e.g. fig. 1 of Latif et al. 2022b). The dense environment favors the continuous accretion by the dominating central SMS, which suppresses the growth of off-center stars (Bonnell et al. 2001; Chon & Omukai 2020). Merger into the SMS is frequent due to its central position and large size (e.g. Stacy et al. 2016; Chon & Omukai 2020), which can also increase its mass. The IMF of the stars formed in the SGC

is thus expected to be significantly affected by these processes and to evolve to be top-heavy. If additional ‘heating’ sources, such as the LW radiation and the turbulence generated by the fast accretion, are present, the formation of small-scale gas clumps can be further suppressed, making the IMF even more top-heavy.

As discussed in §3.1.5, although a gas cloud hosting Pop-III stars is expected to be free of SN feedback, radiative feedback from young Pop-III stars can still present. Recent hydrodynamic simulations (e.g. [Latif et al. 2021, 2022b](#)) incorporating this source of feedback have found that it can significantly regulate the accretion rate of gas into the core of the SGC, and thus limit the total stellar mass of the star cluster formed there. The gas within the SGC was found to be highly dynamical, where gas inflow and feedback-driven outflow co-exist with comparable rates (see e.g. fig. 6 of [Latif et al. 2021](#)). Such strong feedback can ‘quench’ the star formation continuously and lead to an upper limit on the total amounts of stars that can form (see e.g. §3 of [Latif et al. 2022b](#)). Following the same strategy as that adopted by [Paper-I](#) in modeling the AGN-feedback effects on SMBH formation, and by [Paper-III](#) in modeling the stellar-feedback effects on globular-cluster formation, we relate the stellar mass of a Pop-III star cluster to the properties of its host SGC by a balance between the feedback energy/momentum and the SGC binding energy/force, with model parameters appropriate for the Pop-III regime. The top-heavy IMF, together with the fact that $L \propto M_*^2$ for massive stars concerned here, implies that the stellar feedback is dominated by the most massive star, referred to as the first (1st) star or dominant star in the following.

As one extreme case, we first consider a halo at the H₂-cooling threshold, M_{v,H_2} (Eq. 14). The 1st star in such a halo cannot grow too massive, and can approach to the main sequence before the cloud is destroyed (see e.g. fig. 6 of [Stacy et al. 2016](#)). The zero-age main sequence (ZAMS) luminosity is

$$L_* = L_{ZAMS} \approx 1.4 \times 10^4 L_\odot \left(\frac{M_*}{10 M_\odot} \right)^2 \quad (26)$$

(e.g. [Schaller et al. 1992](#); [Hosokawa et al. 2010](#)). The upper limit of the stellar mass, above which the gas accretion is stopped, can be estimated as

$$\frac{\eta_r L_*}{c} = \frac{G M_{SGC}^2}{R_{SGC}^2}, \quad (27)$$

where the free-parameter η_r stands for the coupling efficiency of the stellar radiation to the surrounding gas (and can be > 1 due to multiple scattering and the optically-thick nature of dense SGC). Taking $M_{SGC} = f_b M_{v,H_2}$, and using Eq. (9) for R_{SGC} , we can estimate the star formation efficiency as

$$\begin{aligned} \epsilon_{ZAMS} &\equiv \frac{M_*}{M_{SGC}} = \left[\frac{G(10 M_\odot)^2 c}{(1.4 \times 10^4 L_\odot) \eta_r} \right]^{1/2} R_{SGC}^{-1} \\ &= 2.17 \times 10^{-3} \eta_{r,25}^{-1/2} \left(\frac{R_{SGC}}{3.63 \text{ pc}} \right)^{-1} \\ &= 2.17 \times 10^{-3} \eta_{r,25}^{-1/2} (1+z)_{25}^{3/2}, \end{aligned} \quad (28)$$

where our default value, $\eta_r = 25$, is chosen to be consistent with those obtained by the simulations of [Latif et al. \(2021\)](#), [Latif et al. \(2022b\)](#) and [Latif et al. \(2022a\)](#) for halos of 10^5 – $10^7 M_\odot$ at $z \approx 15$ – 30 . The corresponding stellar mass is

$$M_* = \epsilon_{ZAMS} M_{SGC} = 132 M_\odot \eta_{r,25}^{-1/2}. \quad (29)$$

Interestingly, this value does not depend on redshift, implying that the SGC collapse in a halo with mass at the H₂-cooling threshold

forms a fixed amount of stars. This is a consequence of the fact that the redshift dependencies of ϵ_{ZAMS} and M_{v,H_2} cancel each other.

As the other extreme, we consider an SGC that is supported by strong turbulence in a halo with mass $f_b^{-1} M_{J,max}$ (Eq. 20). In this case, the 1st star has sufficient gas supply to grow massive, with its luminosity limited by the Eddington luminosity,

$$L_* = L_{Edd} \approx 3.284 \times 10^6 L_\odot \left(\frac{M_*}{100 M_\odot} \right) \quad (30)$$

([Hosokawa et al. 2012](#)). Substituting this luminosity into the momentum balance (Eq. 27), taking $M_{SGC} = M_{J,max}$, and using Eq. (11) for Σ_{SGC} , we can estimate the star formation efficiency:

$$\begin{aligned} \epsilon_{Edd} &= \frac{G(100 M_\odot) c}{(3.284 \times 10^6 L_\odot) \eta_r} \Sigma_{SGC} \\ &= 5.29 \times 10^{-3} \eta_{r,25}^{-1} \left(\frac{\Sigma_{SGC}}{2.00 \times 10^4 M_\odot \text{pc}^{-2}} \right) \\ &= 5.29 \times 10^{-3} \eta_{r,25}^{-1} (1+z)_{25}^{3/2}. \end{aligned} \quad (31)$$

This efficiency is only slightly higher than that for a H₂-cooling halo, and is a result of the balance between the higher binding force of the more massive SGC and the stronger radiative feedback of massive stars. The redshift dependence remains the same as that of ϵ_{ZAMS} , and the stellar mass of the 1st star,

$$M_* = \epsilon_{Edd} M_{SGC} = 2.62 \times 10^4 M_\odot \eta_{r,25}^{-1/2}, \quad (32)$$

is again independent of redshift.

Thus, the 1st star in a Pop-III cluster accounts for a few times 10^{-3} of the mass of an SGC at $z \approx 25$, and scales as

$$M_* = \epsilon M_{SGC} \propto M_{SGC}^\beta \quad (33)$$

at a given SGC density, where the star formation efficiency ϵ varies from $\epsilon_{ZAMS} \propto R_{SGC}^{-1}$ (Eq. 28) in pure H₂-cooling halos to $\epsilon_{Edd} \propto \Sigma_{SGC}$ (Eq. 31) in extremely turbulent H-cooling halos. The index, β , thus varies from $2/3$ to $4/3$ between the two extremes. In both cases, the redshift-dependent factors in ϵ and M_{SGC} cancel out, leading to a time-independent M_* - M_{SGC} relation. The delay of the SGC collapse by either the UV radiation or the fast accretion thus affects the mass of the 1st star by elevating M_{SGC} ($\approx f_b M_v$), with the star formation efficiency constrained in a narrow range. Note that we assumed η_r to be the same for the two cases, which is only a crude approximation. In reality, an SMS may be unstable against pulsation, and its radius and temperature can experience episodic oscillations due to a strongly varying inflow rate, making the coupling efficiency, η_r , mass dependent (e.g. [Latif et al. 2022a](#), see their Extended Data fig. 3). The asymmetry in the internal structure of stars and in the distribution of gas around the stars can also affect the coupling efficiency (e.g. [Susa 2013](#); [Latif et al. 2021](#)) and cause halo-to-halo variations in the mass spectrum of Pop-III stars. Such details need to be modeled with numerical simulations.

More general cases are expected to fall between the two extremes described above, and we model them as follows. Once the halo mass for an SGC about to collapse is determined (Eq. 23), we obtain the properties of the SGC using Eqs. (9)–(12), the star formation efficiency $\epsilon = \epsilon_{ZAMS}$ using Eq. (28), and M_* and $L_* = L_{ZAMS}$ of the 1st star using Eq. (30). If L_* so obtained does not exceed L_{Edd} given by M_* , we take M_* as the mass of the 1st star; otherwise, we compute $\epsilon = \epsilon_{Edd}$ from Eq. (31) and use it to find the mass of the 1st star.

Once the mass of the 1st star, $M_{*,1st}$, is determined through the radiation-regulated processes described above, we obtain the full

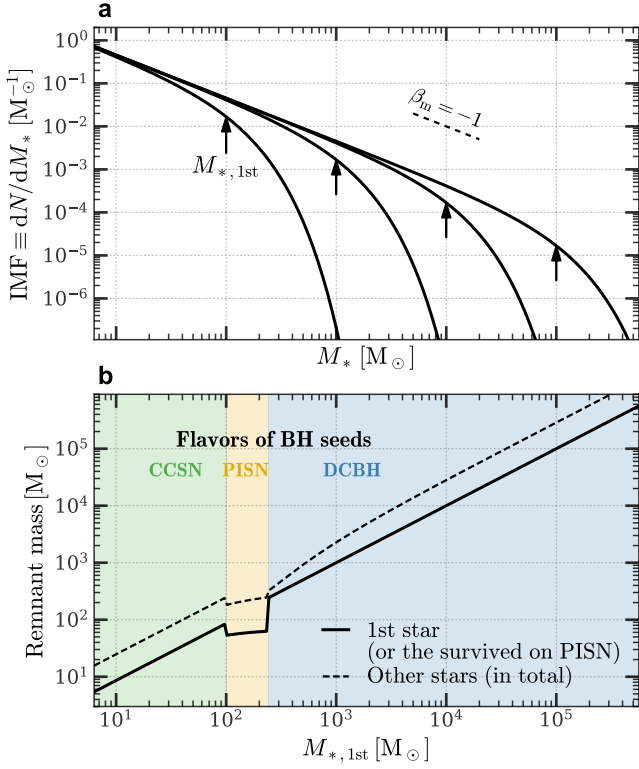


Figure 4. Masses of Pop-III stars and BH remnants. **a**, initial mass function (IMF) of stars in a Pop-III star cluster, each assuming a different mass of the 1st star ($M_{*,1st}$, indicated by the arrow). **b**, remnant mass of stars in a Pop-III star cluster as a function of $M_{*,1st}$ of the cluster, **solid** for the 1st star (i.e. the mass of BH seed) and **dashed** for the other stars in the cluster in total. The evolution of the 1st star (CCSN, PISN or DCBH) defines the flavor of the BH seed. If PISN is triggered to end the life of the 1st star, we take the remnant mass of the most massive star surviving PISN as a replacement. See §3.2 for the details.

IMF using the parametrization:

$$\frac{dN}{dM_*} = \phi_m \left(\frac{M_*}{M_{*,1st}} \right)^{-\beta_m} \exp\left(-\frac{M_*}{M_{*,1st}}\right). \quad (34)$$

For cases without UV heating and with a typical halo accretion rate, cosmological zoom-in simulations of mini-halos performed by Stacy et al. (2016) and Latif et al. (2022b) taking into account radiative feedback from stars give $\beta_m \approx 1.2$. The simulation of Chon & Omukai (2020) assuming an intense LW background gives a similar value, $\beta_m \approx 1$, although the value of $M_{*,1st}$ is much higher. We thus use $\beta_m = 1$ as our default choice, which is also consistent with the choice of Liu et al. (2024) in their modeling of binary black holes. However, we note that the IMF of Pop-III stars is still poorly understood, and other values of β_m are possible (see e.g. Costa et al. 2023, for a number of alternatives). The normalization factor ϕ_m is determined by the constraint that the expected number of stars above $M_{*,1st}$ is unity:

$$1 = \int_{M_{*,1st}}^{\infty} \frac{dN}{dM_*} dM_*, \quad (35)$$

which gives $\phi_m = 4.56/M_{*,1st}$ for $\beta_m = 1$. Thus, at a given β_m , the only parameter needed to fully specify the IMF is $M_{*,1st}$, which we have already determined above from SGC properties and the feedback-regulated scenario. Fig. 4a shows the IMF of stars in a Pop-III star cluster with the default choice of β_m for a number of different

$M_{*,1st}$ values in the range of 10^2 – $10^5 M_{\odot}$. At $M_* \ll M_{*,1st}$, all IMFs converge to the same logarithmic slope of -1 , specified by β_m , and the same normalization. A higher value of $M_{*,1st}$ gives a higher number of massive stars other than the 1st star. As we will see below, massive stars are expected to evolve along pathways different from the less massive ones, thus leaving BH seeds with different flavors.

The total stellar mass other than the 1st star is determined by

$$M_{*,other} = \int_{M_{*,min}}^{M_{*,1st}} M_* \frac{dN}{dM_*} dM_*, \quad (36)$$

where $M_{*,min}$ is the lower limit of the IMF. As shown in Fig. 3, a LW background of $J_{LW,21} \gtrsim 10$ is able to delay the collapse of the SGC to the atomic-cooling regime. An even higher value of $J_{LW,21}$ does not cause further delay, as the atomic cooling is effective to compensate the heating. However, in metal-free gas, a higher $J_{LW,21}$ can prevent the formation of H_2 and can make the collapse closer to an isothermal contraction dominated by Lyman- α cooling. This can generate a much higher inflow rate, promoting the formation of the SMS and suppressing the formation of low-mass stars (Wise et al. 2019; Latif et al. 2021). A similar argument applies to cases where the rapid halo assembly can sustain significant amounts of turbulence to erase small clumps (Latif et al. 2022a). These factors can affect $M_{*,min}$, but the exact value appropriate for it is unknown because of the limited resolution of these simulations. Indeed, as discussed by, e.g. Stacy et al. (2016, see their §6.4), their simulation with a higher resolution can produce stars down to a lower $M_{*,min}$ and predicts an IMF that does not show a peak at $M_* \lesssim 1 M_{\odot}$. In contrast, Chon & Omukai (2020) found the presence of a peak, with the peak mass increasing with the Jeans mass. Despite the uncertainty, the total mass, $M_{*,other}$, in stars other than the 1st is very insensitive to $M_{*,min}$. Using $M_{*,min} = 0$, the integration of Eq. (36) gives

$$M_{*,other} = 2.88 M_{*,1st}, \quad (37)$$

and the variation is less than $0.01 M_{*,1st}$ for any $M_{*,min} < 0.01 M_{*,1st}$.

The 1st star in a Pop-III star cluster is expected to die on a timescale of a few Myr and collapse into a BH (see e.g. §2.2 of Costa et al. 2023). As dynamical interactions among stars tend to sink the more massive object into the center, the BH originated from the 1st star is expected to become the central BH and to grow further by accreting gas and stars from its neighborhood, as to be detailed in §4. The initial mass of the central BH is thus determined by the remaining mass of the 1st star after it dies. Here we adopt the result of Costa et al. (2023, see their §3.1) who used the PARSEC stellar evolution code (Iorio et al. 2023) to follow the evolution tracks of Pop-III stars in different regimes defined by the initial stellar mass, M_* . For a massive star with $M_* \geq M_{*,DCBH} \equiv 242 M_{\odot}$, the remnant is a DCBH with mass $M_{BH} = M_*$. The equality holds as stellar wind is ineffective due to low line-driven opacity originated from the low metallicity (Vink et al. 2001; Baraffe et al. 2001), and the negligible mass loss due to pulsational instability compared with the large accretion rate (see §3.2 Hosokawa et al. 2013). With $100 M_{\odot} \equiv M_{*,PISN} \leq M_* < M_{*,DCBH}$, the pair-instability supernova (PISN) is the trigger to end the life of the star, which totally disrupts the star and leaves no remnant behind. For $M_* < M_{*,DCBH}$, core-collapse supernova (CCSN) is the trigger to end the stellar evolution. We crudely approximate the remnant mass as $M_{BH} = f_{rem} M_*$, with $f_{rem} = 0.86$ to match the numerical result of Costa et al. (2023). A remnant below $3 M_{\odot}$ is classified as a neutron star, but is not expected owing to the large mass of the 1st star.

In the case where the PISN is triggered in the 1st star, we seed a central BH with a mass $M_{BH} = f_{rem} M_{*,surv}$, where $M_{*,surv}$ is the initial mass of the most massive star that can survive the PISN,

obtained by solving

$$1 = \int_{M_{*,\text{surv}}}^{M_{*,\text{PISN}}} \frac{dN}{dM_*} dM_* . \quad (38)$$

The total mass of other stars able to survive is obtained using equations similar to Eqs. (36) and (37), but with the upper limit of integration replaced by $M_{*,\text{surv}}$. In our numerical implementation, to be outlined in §5.1, the mass loss due to stellar evolution for stars other than the 1st is also included when deriving the mass of their remnants. Note that our treatment of the stellar evolution is a simplification. For example, pulsational pair instability (PPI) can take place at a mass close to $M_{*,\text{PISN}}$ and cause fluctuation in the remnant mass (e.g. Farmer et al. 2019; Mapelli et al. 2020; Costa et al. 2023). Star-to-star variation is also expected, due to the dynamical nature of the stellar structure and the dense environment of the host star cluster that leads to frequent interactions among stars.

Fig. 4b shows the remnant mass, separately for the most massive star (1st star at the birth of the star cluster, or the most massive star that can survive the PISN) and other stars in total, as a function of $M_{*,1\text{st}}$ of the Pop-III star cluster. The remnant mass of the most massive star, i.e. the mass of the BH seed, follows closely $M_{*,1\text{st}}$, owing to the insignificant mass loss (zero for DCBH, and $1 - f_{\text{rem}} \approx 0.14$ for CCSN). The exception is for the case of PISN, where the 1st star at birth of the cluster is destroyed and the BH seed is bred by the remnant of the most massive star surviving the PISN. Total remnant mass of other stars also closely follows $M_{*,1\text{st}}$, except for the case of PISN, due to the simple relation between $M_{*,\text{other}}$ and $M_{*,1\text{st}}$ (Eq. 37).

The above discussion for the formation of SGCs, Pop-III star clusters and BH seeds assumes a pristine gas environment in which the SGCs reside. Such a condition is expected only if the IGM is not significantly enriched by previous generations of stars. Since our model can predict star formation of individual galaxies in a cosmological volume, the metallicity of the IGM can also be self-consistently predicted and used to determine when and where the formation of Pop-III stars is possible or prohibited. We present the details for the modeling of IGM enrichment in Appendix A6. Briefly, we find that at $z = 10$, most of the cosmic volume has already been filled by SN bubbles that carry metals produced by the star formation, and the median IGM metallicity goes above $Z_{\text{IGM}} \approx 10^{-4} Z_{\odot}$. This means that halos at $z \gtrsim 10$ form their first generation of stars mainly in the form of Pop-III, and are expected to seed their central BHs according to the model described above. The rapid star formation in galaxies during their fast phases then quickly enriches the IGM, brings the cosmic star formation into the Pop-II regime, and ends the formation of Pop-III stars at $z \approx 10$. In this paper, we adopt a threshold value of $Z_{\text{IGM,Pop-III}} = 10^{-4} Z_{\odot}$ for the formation of Pop-III stars (Bromm & Loeb 2003; Valiante et al. 2011; Costa et al. 2023; Spinoso et al. 2023), and assume that halos embedded in the IGM with metallicity above this value form their first generation of stars in the form of Pop-II. The mass of BH seeds in these halos is expected to be low, assuming that Pop-II stars do not have a top-heavy IMF. We manually set the mass of BH seeds in these halos to be $M_{\text{BH}} = 10 M_{\odot}$, a typical value of the remnant mass of massive stars in normal star clusters.

The enrichment of the IGM by the outflow of metal produced by star formation, and its effect on the formation of stars and BH seeds in nearby halos, are the earliest ‘matter feedback’ of galaxy formation to its environment. This, together with the ‘radiation feedback’ described in §3.1.3, is represented by the orange markers in Fig. 2.

The branch in Fig. 8b summarizes the seeding procedure in the model pipeline. This procedure includes a list of key conditions regarding the competition between cooling and heating to be checked

for the SGC in each halo. These include (i) whether a seed has already been bred in the halo, and thus whether the SGC is self-enriched to cool effectively; (ii) whether the IGM metallicity is high enough so that the SGC is externally enriched to cool effectively; (iii) whether the SGC is exposed to a strong, external UV background that stalls the cooling and delays the SGC collapse; (iv) whether dynamical heating is strong enough to prevent the collapse and fragmentation of the SGC. The fate of the SGC is then determined by the outcome of these conditions, and the BH seed is bred accordingly. Phenomenologically, condition (i) represents the internal effect of galaxy evolution on itself, while conditions (ii)–(iv) are external effects, from either halo assembly or large-scale environment, on galaxy evolution. Such conceptual separation of ‘nature’ and ‘nurture’ has been built for galaxy formation at low z (Peng et al. 2010; Niemi 2011; Casado et al. 2015; Paulino-Afonso et al. 2019; McGibbon & Khochfar 2022; Mucesh et al. 2024; Shi et al. 2024c), while our scenario suggests that all these effects have already been present ubiquitously in the early Universe at $z \gtrsim 20$. Intriguingly, the recent observations of JWST have pushed the discoveries of galaxies and the measurements of environment towards this high- z regime (e.g. Puskás et al. 2025; Shuntov et al. 2025; Lin et al. 2025; Arita et al. 2025), providing opportunities to test our predictions.

4 THE GROWTH FROM SEEDS TO SUPERMASSIVE BLACK HOLES

During the formation of a BH seed, gas can be temporarily expelled from the host SGC by the radiation and potentially also the SN explosion from Pop-III stars. This loss of gas fuel may lead to a quiescent phase in the star formation and AGN activity. Such a phase is expected to be transient, since rapid replenishment of gas is expected as a combined consequence of fast halo assembly ($\gamma_{\text{v}} \gtrsim 1$, see Fig. 3) at the typical redshift of seed breeding ($z \gtrsim 10$) and efficient atomic cooling in mini-halos. The subsequent evolution of the BH seed is thus expected to follow the two-phase scenario as detailed in Paper-I (see their fig. 7 for the growth paths of individual BHs). The critical condition for the growth of a BH is to break the barrier of angular momentum (denoted by J , or by j for the specific value per unit mass) so that gas with low j can reach the zone of influence of the BH within which the gravity is dominated by the BH. Such a condition is naturally achieved in an SGC owing to the early fragmentation that leads to the ballistic motion of sub-clouds (see also Hobbs et al. 2011; Gaspari et al. 2015).

Towards the galactic nucleus, however, the situation is expected to be different. Here we extend the scenario of Paper-I by considering the small-scale processes that can affect the growth of the BH. The key component of the extension is that related to galactic nucleus, including

- (i) the formation of gaseous nucleus that is largely free of supernovae owing to the high density and short dynamical timescale;
- (ii) the turbulence-modified Bondi accretion that is highly efficient in driving gas inflow in the nucleus towards the BH;
- (iii) the formation of turbulent and magnetized accretion disk around the BH due to the rapid inflow;
- (iv) the two modes of accretion, one ‘wet’ and the other ‘dry’, that can reach a super-Eddington rate and cause the BH to grow fast in short bursts.

The numerical model for the nuclear processes is then built upon these considerations and implemented as a refinement engine of the whole growing procedure so that BH growth can be resolved self-consistently within the context of the cosmic hierarchy. Fig. 5

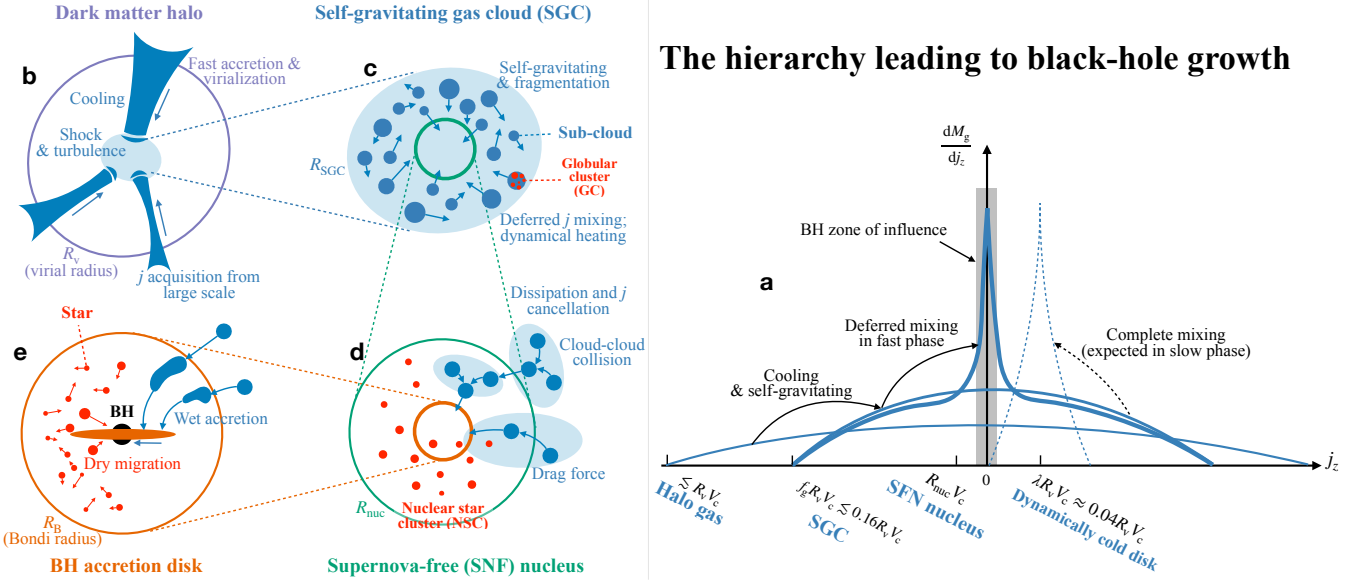


Figure 5. Schematic diagram showing the cosmic hierarchy that leads to black-hole growth in a fast-accreting halo. **a**, the distributions of gas specific angular momentum (j_z , along the direction of net gas spin) shaped by the gas processes in the cosmic hierarchy. **b–d**, processes in each level of the hierarchy. The representations of the values are not to scale, but to show their relative sizes. Gas accreted together with fast assembly of halo, likely in the form of filaments, rapidly cools and flows into the halo center (**b**). The j_z distribution of inflow gas is initially broad, depending on the accretion configuration, and becomes narrower due to radiative cooling. Inflow gas collides to generate strong shocks, yields supersonic turbulence, and becomes self-gravitating before angular-momentum support kicks in, thus forming a self-gravitating, turbulent gas cloud (**c**; SGC). Rich gas and frequent perturbations cause the cloud to fragment into sub-clouds, whose density is further elevated by shock compression, with some being able to form globular clusters. Dissipation that mixes (cancels) the angular momentum of the sub-clouds is deferred due to infrequent cloud-cloud collision and weak drag force by surrounding gas, preventing the formation of a dynamically-cold disk and leaving the j_z distribution broad. Under certain conditions (e.g. global disturbances) that drive strong gas inflow towards galactic center, a high-density, supernova-free (SNF) nucleus is expected to form (**d**). Angular-momentum mixing is thus activated surrounding and within the nucleus (**d**), restructuring j_z to form a centralized and peaked distribution, and routing the gas towards the central BH. Within the zone of influence of the BH (**e**), gas surviving the AGN feedback joins into the turbulent and magnetized accretion disk and feeds the BH through this ‘wet channel’. Some stars formed in the nucleus migrate inwards via dynamical relaxation, and feed the BH through this ‘dry channel’. The mixing of angular momentum at SGC scale and the build-up of such a barrier of BH accretion (dashed distribution in **a**) are expected to be efficient only in the slow phase. See §4 for the details.

previews the four levels of the hierarchy from the dark matter halo to the BH accretion disk, and processes that can redistribute j . Fig. 6 shows the gas density profile at each level, at three redshifts, predicted for the BHs experiencing super-Eddington accretion in an MW-size galaxy and its progenitors experiencing fast assembly. In the following, we discuss them in details.

4.1 The hierarchy leading to black-hole growth

The outermost boundary in the hierarchy of the cosmic structure formation is the Universe itself, given by a Λ CDM cosmology with a fiducial set of parameters, and the large-scale structure formed by the gravitational instability in the Universe. The formation of virialized, dark-matter-dominated halos causes a jump in the density to a level about $\Delta_\nu \sim 200$ times the cosmic mean density, as detailed in §2. The gas density associated with the dark matter halo is also lifted to be about Δ_ν times the cosmic mean baryon density, $\rho_b(z)$ (Eq. 1). At such density, the cooling of the gas is significantly enhanced due to the n^2 dependence of the two-body cooling processes, bringing the time needed for the condensation of the gas well below the Hubble timescale, $H^{-1}(z)$ (see e.g. the cooling diagram in fig. 2 of Paper-I). The formation of halos thus forms a necessary step for the formation of stars and the growth of BHs in galaxies.

Fig. 6a shows the gas density as a function of the halo-centric distance at three epochs in the progenitor of an MW-size halo. The density profile exhibits a ladder-like structure, consisting of five steps

that correspond to the five levels of the cosmic hierarchy. Each purple segment shows the gas density at virialization of the halo, assuming an isothermal profile with temperature equal to T_ν (Eq. 5; see also Croton et al. 2006; Guo et al. 2011; Lacey et al. 2016). The card in Fig. 6b summarizes equations governing halo properties, as presented in §2.

A key property of CDM halos, as mentioned in §1, is their two-phase assembly. The early evolution of a dark matter halo is thus expected to consist of processes distinct from those later on. These include:

- (i) fast accretion, with $\gamma_\nu \gtrsim \gamma_f$, where the threshold value $\gamma_f \approx 3/8 - 0$ (Zhao et al. 2003; More et al. 2015; Boco et al. 2023; see also §2 of Paper-I for a summary);
- (ii) efficient cooling, with $t_{\text{cool}} \lesssim t_\nu$, if halo mass has not grown above a redshift-independent threshold of $\approx 10^{12} M_\odot$ (see e.g. fig. 2 of Paper-I);
- (iii) cold stream in halo gas, depending on the configuration of cosmic filaments connected to the halo (Keres et al. 2005; Dekel et al. 2009; Ceverino et al. 2010; Danovich et al. 2015).

Processes (i) and (ii) in combination are sufficient to maintain rapid inflow and keep the halo center gas-rich before feedback can eject much of the gas. The high gas fraction then implies that the gas can become self-gravitating when the inflow reaches a radius of $f_b R_\nu \lesssim f_b R_\nu \approx 0.16 R_\nu$ before angular momentum becomes important to resist the collapse (at a scale of $\lambda_g R_\nu \approx 0.04 R_\nu$, assuming that the spin of the gas, λ_g , follows that of the dark matter). This leads to the

Density ladder of the hierarchy

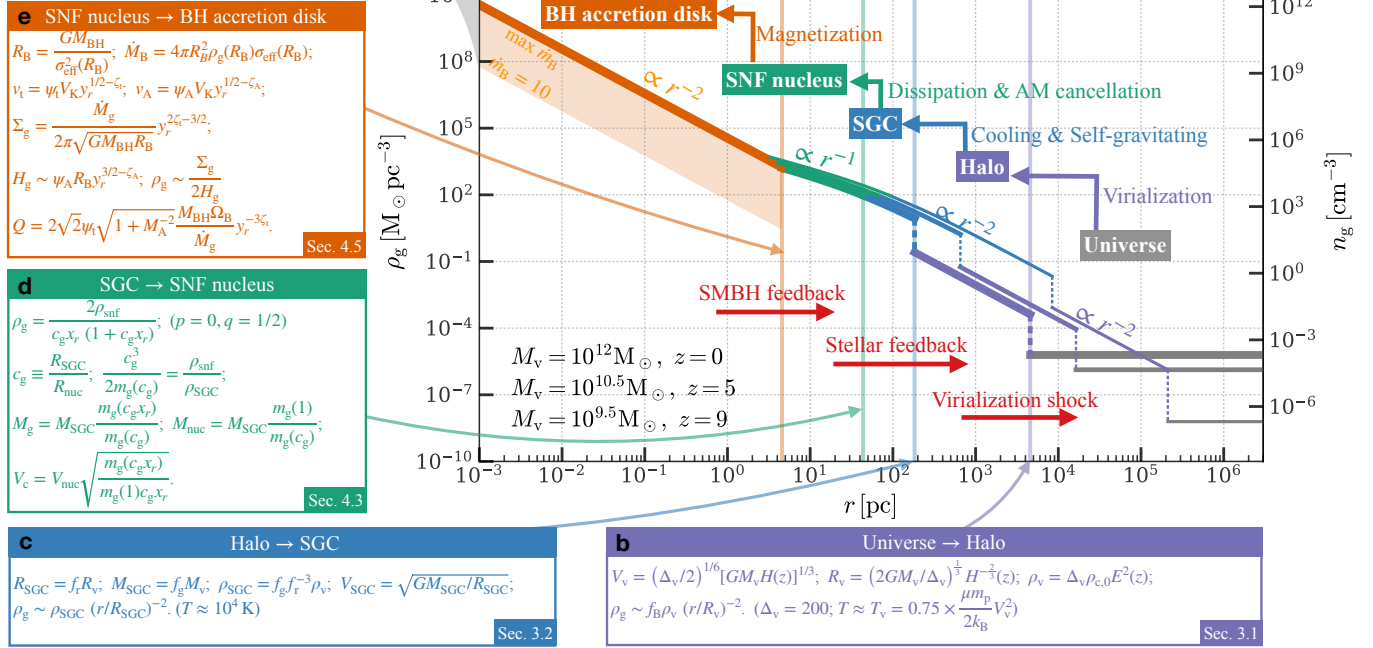


Figure 6. Density ladder of gas in the cosmic hierarchy. **a**, gas density profiles predicted by our model at three redshifts, $z = 0, 5$ and 9 (from thin to thick), in and around an MW-size galaxy and its progenitors under global disturbance. At each redshift, an SMBH with $M_{\text{BH}} = 10^6 M_{\odot}$ is put at the center of the galaxy. The profile exhibits a ladder-like structure, separated into **five segments**, each corresponding to a level in the structure hierarchy discussed in this paper. The innermost segment (**orange**) describes the middle-plane density of the accretion disk around the central BH, with a solid line indicating the profile expected in the beginning of the nuclear burst when the turbulence-modified Bondi accretion rate, \dot{m}_{BH} (in the unit of Eddington rate), is the maximum (Eq. 69), and a shaded area covering the range of the profile during the burst until \dot{m}_{BH} is reduced to 10. Each of the four **vertical lines** indicates the transition scale of the interface bridging two adjacent levels at $z = 9$. **b–d**, cards listing the sets of equations bridging the structure properties of any outer level to those of the adjacent inner level. The processes driving the transitions, and the feedbacks generated from different levels, are indicated by the texts. See §4.1 for details.

formation of a self-gravitating gas cloud in the halo center, as shown schematically in Fig. 5b. The motion of gas elements in the SGC at the time of self-gravitation is thus dominated by random motion instead of rotation, and the dispersion of j at an arbitrary direction is about $f_g R_v V_c$, where V_c is the circular velocity at the SGC scale (see Fig. 5a).

Process (ii) alone implies that the gas within the halo can fall freely towards the halo center, with a velocity of $\approx V_v$, without affected significantly by thermal pressure. Shock is thus expected to be produced when the high-flux, high-speed inflow collides within the SGC. This, together with the frequent dynamical disturbance imparted by the rapid change of the gravitational potential during the fast-accretion regime, can generate supersonic turbulence, with the Mach number $\mathcal{M}_s \equiv v_t/c_s \gg 1$, over the entire SGC. Fragmentation in the turbulent medium due to Jeans instability and shock compression produces sub-clouds with densities as high as $n_{\text{sc}} \approx \mathcal{M}_s^2 n_{\text{SGC}}$, making both the cloud-cloud collision and the drag force from the surrounding gas negligible. The presence of process (iii) can further enhance the turbulence by increasing the fraction of kinetic energy that can penetrate the SGC, which can accelerate the formation of sub-clouds by promoting early fragmentation of gas in the filamentary inflow. As soon as the collision and drag become unimportant, the mixing of angular momentum among sub-clouds ceases, leaving a set of ballistically moving sub-clouds that support the structure of the SGC. Sub-clouds with highest density ($n_{\text{sc}} \gtrsim 10^{3.5} \text{ cm}^{-3}$) are predicted to be the sites to form globular clusters, as modeled

in Paper-III. Fig. 5c is a schematic plot showing the structure of an SGC, formed in a halo (panel b), together with its sub-clouds. As a consequence of the formation of sub-clouds, the broad distribution of j on the SGC scale is preserved, in contrast to the peaked, off-center distribution that would be expected for a dynamically-cold disk formed in a halo in the slow phase, as indicated in Fig. 5a. A fraction of the sub-clouds with low j in the broad distribution is thus expected to reach the influence zone of, and to be captured by, the central BH, as indicated by the vertical grey band in Fig. 5a.

The formation of the SGC produces the jump in density from the halo density produced by halo formation. Blue segments in Fig. 6a show the resulted density profiles at this step, where we have assumed an isothermal density profile (Henriques et al. 2020; Hopkins et al. 2024a) with a temperature, $T \approx 10^4 \text{ K}$, at which a cooling barrier is present (see e.g. the cooling functions in fig. 4 of Maio et al. 2007 and fig. 1 of Smith et al. 2008). The blue card in Fig. 6c summarizes equations relevant to this step, as presented in §3.1.1.

Towards the galactic nucleus, several additional processes need to be considered. The first is related to the dominant role of the BH in regulating the amount of gas that can be retained in the nucleus. Such a regulation is clearly seen in hydro simulations that include recipes for AGN feedback (e.g. Sivasankaran et al. 2025) as a reduction in the BH accretion rate produced by the feedback from the BH itself. The second is related to the regeneration of low- j sub-clouds that would otherwise be absent after the BH has captured them all or destroyed them via feedback (Sivasankaran et al. 2025, see their fig. 5). In Paper-

I, we speculated that the ‘positive’ SN feedback can play a role in this regeneration, as is suggested by the findings of hydro simulations that nuclear star formation is not passive (e.g. [Bower et al. 2017](#); [Habouzit et al. 2021](#)). Stellar feedback in the nucleus with high density may not be able to unbind the gas (e.g. [Dekel et al. 2023](#); [Hopkins et al. 2024a](#)), but can drive turbulence (e.g. [Ma et al. 2020](#)) and re-create sub-clouds to feed the BH. The third is related to the mixing of angular momentum in the galactic nucleus, as the gas density involved can be high enough to generate significant drag force and frequent cloud-cloud collisions. This has been clearly seen in the simulation of [Gaspari et al. \(2015, see their fig. 15\)](#), in which even weak turbulence promotes the redistribution of angular momentum in galactic nucleus and the BH feeding is expected to be replenished by the low angular-momentum sub-clouds generated by the redistribution. The results of all these processes, however, depend on the competition between gas heating and cooling. For example, SN feedback, cloud-cloud collision, drag force and AGN feedback can inject energy into the gas, which may destroy sub-clouds if cooling is ineffective, instead of promoting sub-cloud formation. Thus, these additional processes may cause the gas structure and dynamics to deviate from those expected from modeling the outer part of the SGC, thereby modifying the density ladder, angular momentum distribution and BH accretion rate. We therefore model the extension of the hierarchy to small scales in the following.

4.2 The formation of supernova-free nuclei

The aforementioned processes within a galactic nucleus require an additional refinement on top of the SGC to capture their consequences for gas dynamics, star formation and BH accretion. To avoid ambiguity, we define the nucleus as the inner part of the SGC within which gas density is above a ‘supernova-free’ (SNF) threshold,

$$n_{\text{snf}} = 10^{3.5} \text{ cm}^{-3}, \quad (39)$$

or

$$\rho_{\text{snf}} = \mu m_{\text{p}} n_{\text{snf}} \approx 6.35 \times 10^{-21} \text{ g cm}^{-3}. \quad (40)$$

Using Eq. (13), the free-fall timescale of the gas at a density $\rho_{\text{g}} = \mu m_{\text{p}} n_{\text{g}}$ is

$$t_{\text{ff,g}} = \sqrt{\frac{3\pi}{32G\rho_{\text{g}}}} = [0.836 \text{ Myr}] n_{\text{g},3.5}^{-1/2}, \quad (41)$$

where $n_{\text{g}} \equiv 10^{3.5} \text{ cm}^{-3} n_{\text{g},3.5}$. Thus, a density of $n_{\text{g}} \geq n_{\text{snf}}$ corresponds to a free-fall timescale of $t_{\text{ff,g}} \leq t_{\text{snf}} \equiv 0.836 \text{ Myr}$, implying that gas dynamics within the nucleus proceeds faster than the lifetime of massive stars born in it. Such a nucleus is thus largely free of supernovae. In the following, we refer to this volume as the **SNF nucleus**. Another direct consequence of the high density and short dynamical timescale is that stellar feedback, including stellar winds, radiation and supernovae, is ineffective in dispersing the gas, as all the feedback effects require the presence of evolved massive stars. As detailed in [Dekel et al. \(2023\)](#) and [Paper-III](#), the dense gas in the SNF regime may also shield itself against external effects, such as evaporation by an external radiation field. The two potential ways to remove gas from an SNF nucleus are (i) the consumption of the gas by star formation/BH; (ii) the removal of the gas by the feedback from the accreting BH. These will lead to changes in the structure of the nucleus, by forming an extremely compact stellar system with mass density $\sim \rho_{\text{snf}}$ and by the rapid growth of the BH. Both of the mass components (the compact stellar cluster and the BH) are expected to build their mass in a highly bursty fashion, with each

burst lasting for a period of $\leq 1 \text{ Myr}$. In what follows, we refer to such bursts in star formation and BH growth as **nuclear bursts**, and we will model them in §4.5. Fig. 5d shows the sketch of the SNF nucleus as a zoom-in view of the central part of the host SGC. The relations of the nucleus to the expected nuclear star cluster (NSC) and central BH are also highlighted. In the following, we describe the formation and structure of the SNF nucleus in detail.

4.2.1 Criteria for the formation of supernova-free nuclei

Questions remain as how the SNF nucleus can be formed in the first place. Several processes can present barriers for the formation, mainly,

- (i) the formation of dynamically-cold disk at the SGC scale, which can build an angular-momentum barrier for the gas flow towards the nucleus (e.g. [Power et al. 2011](#); [Rosas-Guevara et al. 2015](#); [Curtis & Sijacki 2016](#));
- (ii) star formation and stellar feedback, which may consume/eject much of the gas before it can reach to the nucleus ([Meng & Gnedin 2020](#); [Renaud et al. 2021](#); [van Dokkelaar et al. 2022, 2023](#); [Garcia et al. 2023](#); [Renaud et al. 2024](#); [Mayer et al. 2024](#));
- (iii) feedback from the accreting BH, which may affect gas accumulation in the nucleus ([Su et al. 2023](#); [Dutta et al. 2024](#); [Sivasankaran et al. 2025](#); [Su et al. 2025](#)) or may even eject/heat the gas in the SGC and halo ([Zinger et al. 2020](#); [Shi et al. 2022](#); [Ayromlou et al. 2023](#); [Weinberger et al. 2023](#); [Wang et al. 2024a](#)).

The formation of SNF nucleus is thus expected only when a set of conditions are met to break these barriers.

In our modeling relevant to the early fast accretion regime of the halo assembly, the creation of a dynamically hot, globally disturbed SGC is essential to remove those barriers. The ‘hotness’ is needed to prevent the formation of dynamically cold disks, so as to break the angular-momentum barrier (point i above). Global disturbance can make an SGC out of equilibrium, which can produce a strong gas inflow to overpower (at least over a short period of time) the gas consumption by star formation/BH accretion and ejection by the associated feedback (points ii and iii above). Once these conditions are met, an SNF nucleus can be formed as a consequence of the boosted inflow, and the formation of a compact stellar system and a period of rapid growth of the central BH seem inevitable.

The conditions for the hotness and global disturbance of an SGC arise naturally owing to the fluctuations in the halo assembly expected in the fast accretion regime (see examples in fig. 1 of [Paper-I](#)). The temporary speed-up of the halo mass assembly implies an excessive amount of matter being added to the halo, which can perturb the SGC significantly through changing the gravitational potential and by the ram pressure of the infalling gas. Part of such disturbance is produced by mergers with other halos, which can perturb the SGC during an close encounter of the merging galaxies, a process known to be critical in transforming galactic structure (e.g. [Springel & Hernquist 2005](#); [Keselman & Nusser 2012](#); [Puech et al. 2012](#); [Yoon et al. 2022](#); [Jackson et al. 2022](#); [Wright et al. 2025](#); [Sampaio et al. 2025](#); [Proctor et al. 2025](#)). A halo temporarily taking an excursion to a sufficiently high mass assembly rate is thus expected to be able to impart global disturbance and create the hotness required for the formation of an SNF nucleus in the SGC. These considerations motivate us to adopt the following criterion to trigger the formation of an SNF nucleus:

$$\gamma_{\text{v}} - \mathcal{S}(\gamma_{\text{v}}; t_{\text{v}}) \geq \delta_{\gamma} \gamma_{\text{f}}, \quad (42)$$

where $\mathcal{S}(\gamma_{\text{v}}; t_{\text{v}})$ denotes the value of γ_{v} smoothed over a timescale of t_{v} along the main branch of the halo assembly history (see Appendix A4 for details), $\gamma_{\text{f}} \approx 0\text{--}3/8$ is the threshold that separates

the fast and slow phases of halo assembly (Paper-I), and $\delta_\gamma \sim 1$ is a parameter that defines the strength of the excursion needed for the formation of an SNF nucleus. Here, we adopt $\gamma_f = 3/16$, the average of the two extremes, and we choose $\delta_\gamma = 1$ to identify excursions that have relaxation time scales longer than the typical dynamical time of the halo.

The frequency of the formation of SNF nuclei is expected to be high in the early Universe. This can be understood as a result of the specific merger rate of CDM halos (Dong et al. 2022) in combination with their faster assembly rate at higher z (e.g. Fig. 1). As we will model in §4.5 and show in §5.3.1, the growth of the BH in an SNF nucleus is expected to be super-Eddington. This, together with the high frequency of formation of SNF nuclei, provides an efficient way for BH seeds to grow into SMBHs.

4.2.2 Structure of the supernova-free nuclei

The violent gas inflow associated with the formation of SNF nuclei creates a situation similar to that for the formation of SGC, albeit with smaller volume, higher density, and shorter timescale. The large mass flux ($\sim M_{\text{SGC}}/t_{\text{dyn,SGC}}$) and high velocity ($\sim V_{\text{SGC}}$) are similar to the conditions of the SGC formation provided by a halo of fast assembly and efficient-cooling, and are also expected to drive supersonic turbulence in the nucleus. Such a gaseous nucleus is thus dynamically hot (with turbulence velocity $v_t \gg c_s$), with structure supported by the turbulent motion. This gives some clue about the structure expected from an SNF nucleus.

However, the gas dynamics within an SNF nucleus is not entirely the same as that in an SGC. One key difference is the lack of regulation by supernova and AGN feedback in the nucleus, since the formation of the nucleus, as detailed in §4.2.1, requires temporarily boosted inflow so that star formation, BH accretion and their feedbacks are incapable of balancing the inflow rate. Instead, the main processes diving energy transfer during the formation of an SNF nucleus are the release of gravitational potential energy carried by the inflow and the dissipation of the turbulent energy by interacting gas elements. The situation here thus resembles that in hydrodynamic simulations of isothermal turbulence with a single driving source, such as that mimicked by a stochastic force term or a shaking box in such simulations (Eswaran & Pope 1988; Schmidt et al. 2009; Federrath et al. 2021). The real gas profiles concerned here may still be different from those in the simulations, as the SNF nuclei may not be able to fully develop an equilibrium, in which energy injection balances dissipation, before nuclear star formation and BH accretion are elevated to disperse the gas. The lack of regulation and the run-away nature of SNF nuclei during formation may thus cause uncertainties in the modeling based on the assumption of equilibrium.

As an approximation, we divide the processes in an SNF nucleus into two stages. The first, referred to as the formation stage, is the period when the gas inflow is boosted to replenish the gas content in the nucleus and the regulation from the feedback has yet to be established. The second, referred to as the dispersion stage, is the period of a nuclear burst, in which star formation and BH accretion are elevated so that the gas consumption and dispersion by them overpower the inflow before the gas density within the nucleus is reduced to the original level. For the formation stage, we will model the structure equations of the gas around and within an SNF nucleus by a parametric approach in the following, with a number of free parameters to carry the unknown degree of freedom. For the dispersion stage, we model the amount of the star formation and BH accretion in the nuclear burst together with the reduction of the gas in §4.5.

We parameterize the gas density (ρ_g) as a function of the galacto-

centric distance (r) within the entire SGC, including that within the SNF nucleus during the formation stage, by the following profile:

$$\rho_g(r) = \frac{2^{p+2q} \rho_{\text{snf}}}{(c_g x_r)^{2-2q} (1 + c_g x_r)^{p+2q}}, \quad (43)$$

where the scaled radius $x_r = r/R_{\text{SGC}}$; the concentration parameter $c_g \equiv R_{\text{SGC}}/R_{\text{nuc}}$; R_{nuc} is the outer radius of the SNF nucleus; ρ_{snf} is the supernova-free threshold (Eq. 40); $q \leq 1$ is the structure parameter of the turbulence (see below); $p \geq 0$ characterizes the density profile at large radius. It can then be shown that $\rho_g(R_{\text{nuc}}) = \rho_{\text{snf}}$. Thus, $\rho_g \geq \rho_{\text{snf}}$ for any $r \leq R_{\text{nuc}}$, consistent with the requirement that the nucleus is supernova-free.

The double power-law form of the profile divides the gas structure of the SGC into two regimes: one at $r \gg R_{\text{nuc}}$, where the density $\rho_g \propto r^{-(2+p)}$, and the other at $r \ll R_{\text{nuc}}$, where the density $\rho_g \propto r^{-(2-2q)}$. This separation is plausibly produced by distinct processes. Specifically, the parameter q can be understood in terms of the velocity structure of a turbulent medium, $\sigma_v \propto r^q$, namely the velocity dispersion σ_v of the gas as a function of the scale r . Assuming virialization, the mass enclosed within r is $M_g(r) \propto r \sigma_v^2 \propto r^{2q+1}$. The density profile is thus $\rho_g \equiv dM_g / (4\pi r^2 dr) \propto r^{-(2-2q)}$, which is the same as our parameterization (Eq. 43) at small r . The mean density within r can then be written as $\bar{\rho}_g(r) = 3\rho_g(r)/(2q+1)$. We choose $q = 1/2$ as the default value. This choice is commonly used for well-relaxed, isothermal and supersonic turbulence (Federrath et al. 2021), and we adopt it because of the supersonic nature of the SNF nuclei expected from the rapid inflow driven by the global disturbance. The choice of q thus implies that Eq. (43) applies to a well-relaxed state of the SNF nucleus, where dissipation by radiative cooling is balanced by the energy injection into the gas. This balance is expected to be transitional, however, as the bursty star formation and BH accretion will soon end the formation stage and take the nucleus into the dispersion stage.

We adopt $p = 0$ as the default value, which corresponds to an isothermal profile, $\rho_g \propto r^{-2}$, for the volume of SGC well outside the nucleus. Such a gas profile is seen in hydro simulations (e.g. Hopkins et al. 2024a), and is adopted by semi-analytical models of galaxy formation (e.g. Henriques et al. 2020). With $p = 0$ and $q = 1/2$ given above, Eq. (43) becomes

$$\rho_g(r) = \frac{2\rho_{\text{snf}}}{c_g x_r (1 + c_g x_r)}, \quad (44)$$

and the corresponding asymptotic behaviors are $\rho_g \propto r^{-2}$, $\bar{\rho}_g(r) = 3\rho_g(r)$, $M_g \propto r$, and $V_c = V_{\text{SGC}} = \text{const.}$ for $r \gg R_{\text{nuc}}$; $\rho_g \propto r^{-1}$, $\bar{\rho}_g(r) = 3\rho_g(r)/2$, $M_g \propto r^2$, and $V_c \propto r^{1/2}$ for $r \ll R_{\text{nuc}}$.

To determine the structure of an SNF nucleus from the global properties of the host SGC (Eqs. 9–13), we can substitute the constraint $M(R_{\text{SGC}}) = M_{\text{SGC}}$ into the profile Eq. (43), and obtain the concentration parameter c_g by solving the following implicit equation:

$$\frac{2}{c_g} \frac{m_g(c_g)}{c_g^2} = \frac{\rho_{\text{SGC}}}{\rho_{\text{snf}}}, \quad (45)$$

where $m_g(x) \equiv 3[x - \ln(1+x)]$. The radius of the SNF nucleus is then given by the definition, $R_{\text{nuc}} = R_{\text{SGC}}/c_g$. The mass enclosed within r is obtained by integrating $4\pi r'^2 \rho_g(r')$ over $r' \in [0, r]$:

$$M_g(r) = M_{\text{SGC}} \frac{m_g(c_g x_r)}{m_g(c_g)}. \quad (46)$$

The mass of the SNF nucleus, $M_{\text{nuc}} \equiv M_g(R_{\text{nuc}})$, is then

$$M_{\text{nuc}} = M_{\text{SGC}} \frac{m_g(1)}{m_g(c_g)}, \quad (47)$$

where $m_g(1) \approx 0.921$, and the circular velocity profile is

$$V_c(r) \equiv \sqrt{\frac{GM_g(r)}{r}} = V_{\text{nuc}} \sqrt{\frac{m_g(c_g x_r)}{m_g(1)c_g x_r}}, \quad (48)$$

with $V_{\text{nuc}} \equiv (GM_{\text{nuc}}/R_{\text{nuc}})^{1/2}$. For $r \ll R_{\text{nuc}}$ (i.e. $c_g x_r \ll 1$), V_c can be approximated as

$$V_c \approx \sqrt{1.5c_g x_r / m_g(1)} V_{\text{nuc}} \approx 1.28 (c_g x_r)^{1/2} V_{\text{nuc}}. \quad (49)$$

To obtain numerical estimates of some quantities in the above equations, we consider $c_g \gg 1$, which holds for SGC at $z \lesssim 10$ (see Fig. B1b). In this case, $m_g(c_g) \approx 3c_g$, and the above equations for the structure of the SNF nucleus reduce to

$$\frac{3}{m_g(1)} \frac{M_{\text{nuc}}}{M_{\text{SGC}}} \approx \frac{R_{\text{nuc}}}{R_{\text{SGC}}} = \frac{1}{c_g} \approx \left(\frac{\rho_{\text{SGC}}}{6\rho_{\text{snf}}} \right)^{1/2}. \quad (50)$$

Taking the estimates of R_{SGC} and n_{SGC} in Eqs. (9) and (10), we obtain

$$c_g^{-1} \approx 0.239 f_{g,0.157}^{1/2} f_{r,0.04}^{-3/2} n_{\text{snf},3.5}^{-1/2} (1+z)_{10}^{3/2}; \quad (51)$$

$$R_{\text{nuc}} \approx [43.6 \text{ pc}] f_{g,0.157}^{1/2} f_{r,0.04}^{-1/2} n_{\text{snf},3.5}^{-1/2} M_{v,9.5}^{1/3} (1+z)_{10}^{1/2}; \quad (52)$$

$$M_{\text{nuc}} \approx \left[6.00 \times 10^7 M_{\odot} \right] f_{g,0.157}^{3/2} f_{r,0.04}^{-3/2} \times n_{\text{snf},3.5}^{-1/2} M_{v,9.5} (1+z)_{10}^{3/2}, \quad (53)$$

where $M_{v,9.5}$ is the halo mass scaled by $10^{9.5} M_{\odot}$, roughly the value of the progenitor at $z = 9$ of an MW-size halo at $z = 0$ ($M_{v,z=0} \approx 10^{12} M_{\odot}$; see Fig. 1). The free-fall timescale and angular frequency of the nucleus are

$$t_{\text{ff,nuc}} = \sqrt{\frac{3\pi}{32G\bar{\rho}_g(R_{\text{nuc}})}} \approx 0.616 \text{ Myr } n_{\text{snf},3.5}^{-1/2} \quad (54)$$

$$\Omega_{\text{nuc}} \equiv R_{\text{nuc}}/V_{\text{nuc}} \approx 1.80 \text{ Myr}^{-1} n_{\text{snf},3.5}^{1/2}, \quad (55)$$

respectively. Using the structure equations of the nucleus (Eq. 50) and the numerical estimate of Σ_{SGC} (Eq. 11), the surface density of the SNF nucleus is

$$\begin{aligned} \Sigma_{\text{nuc}} &\sim \frac{M_{\text{nuc}}}{R_{\text{nuc}}^2} \approx \frac{m_g(1)}{3} c_g \Sigma_{\text{SGC}} \\ &\approx 3.15 \times 10^4 M_{\odot} \text{pc}^{-2} [\chi_g(1+z)_{10}]^{1/2} M_{v,9.5}^{1/3}, \end{aligned} \quad (56)$$

where $\chi_g = f_{g,0.157} n_{\text{snf},3.5} f_{r,0.04}^{-1}$. The circular velocity of the nucleus, V_{nuc} , is related to that of the SGC, V_{SGC} , by $V_{\text{nuc}} = \sqrt{m_g(1)/3} V_{\text{SGC}}$.

Green segments in Fig. 6a extend the density profiles to the nuclear volumes of SGCs in MW-like progenitors by assuming that SNF nuclei have formed. For reference, in Appendix B, we show the properties of SNF nuclei in halos with a wide range of mass over an extended range of redshift. The two pieces of each SGC profile at $r \ll R_{\text{nuc}}$ and $r \gg R_{\text{nuc}}$ are joined smoothly by the function given by Eq. (44). The inner and outer slopes (in logarithmic scale) are -1 and -2, respectively, as an outcome of our assumption on the structure parameters, p and q . The nucleus size, $R_{\text{nuc}} \lesssim 100 \text{ pc}$, shows only slow evolution from $z = 9$ to $z = 0$, much slower than the evolution of R_{SGC} of the host SGC (and R_v of the host halo). This is due to the requirement of a constantly high density, $\rho_g(R_{\text{nuc}}) = \rho_{\text{snf}}$, and the cancellation of two effects (see Eq. 50): the growth of R_{SGC} (and

R_v) with galaxy (and halo) assembly, and the decline of ρ_{SGC} (and ρ_v) with the cosmic expansion. The nucleus size is roughly 1/4 of the SGC size at $z = 9$ (1/80 at $z = 0$), corresponding to 1/100 (1/2000) of the halo virial radius, respectively. This indicates that the SNF nucleus constitutes a significant part of the host SGC at high z , but becomes only a tiny part at low z . This also implies that SNF nuclei play the dominating role in supplying BH growth at high z , and a negligible role at low z , as we will discuss later in §5.3.1 with Figs. 13 and 14. The card in Fig. 6d summarizes equations that link the properties of the SNF nucleus with those of the host SGC.

An interesting conclusion can be reached by considering the time evolution of structural properties of the SNF nucleus in an SGC. This can be quantified by taking the time derivative of, e.g. Σ_{nuc} (Eq. 56), along the main branch of a halo:

$$\frac{d \ln \Sigma_{\text{nuc}}}{d \ln a} \approx -\frac{1}{2} + \frac{1}{3} \frac{d \ln M_v}{d \ln a} = \frac{2\Gamma_v - 3}{6}, \quad (57)$$

where $a = 1/(1+z)$, and $\Gamma_v \equiv \dot{M}_v/(M_v H)$ is the specific growth rate of the halo mass. Note that this equation does not describe the evolution of a single nucleus; instead it describes the difference between nuclei formed at different epochs in the main branch of a halo. The derivative of Σ_{nuc} changes its sign at a critical value of $\Gamma_v = 1.5$, coinciding with the epoch found by Boco et al. (2023) that separates the fast- and slow-accretion phases of a halo (equivalent to $\gamma_v = 3/8$, our ‘Low- z_f ’ variant defined in §2 of Paper-I). The same arguments applied to R_{nuc} and M_{nuc} using Eqs. (52) and (53), respectively, yield the same critical value of Γ_v . Note that the presence of a critical value of Γ_v , and its exact value of 1.5, are owing to the assumption of the isothermal profile in the outer SGC (and thus the index of 1/2 in Eq. 50) and independent of the assumption made for the profile at $r \lesssim R_{\text{nuc}}$. The drag term, $-1/2$, in Eq. (57) comes from the decrease of the density of a virialized halo with the cosmic time, and thus encodes the effect of cosmological expansion on the growth of structures deeply embedded in centers of dark matter halos.

As we will show later in §4.3, the turbulent motion of gas within the nucleus can induce strong inflow toward the central BH, and both the Bondi radius (R_B ; Eq. 68) and the Bondi accretion rate (\dot{m}_B ; Eq. 69) of the BH are determined by Σ_{nuc} . The structure of the accretion disk of the BH, as to be quantified in §4.4 by the profiles of the surface density (Σ_g ; Eq. 81), the volume density (ρ_g ; Eq. 82), and the Toomre Q parameter (Eq. 83), all depend on Σ_{nuc} through the Bondi properties. Hence, the transition of the halo assembly also leads to a transition in BH accretion properties. This is a highly non-trivial result of our model, as it links the cosmological assembly of the structure on halo scales ($r \gtrsim 100 \text{ kpc}$ for $M_v \gtrsim 10^{11} M_{\odot}$ at $z = 0$) to that close to the central BH ($r \lesssim$ a few pc for $M_{\text{BH}} \lesssim 10^6 M_{\odot}$).

The mass scale of an SNF nucleus, M_{nuc} , is typically a few times $10^7 M_{\odot}$ (Eq. 53), about one order of magnitude larger than that of a sub-cloud at a comparable density (see Eq. 29 of Paper-III). The mass scale is also much larger than the typical mass of BH seeds, including massive DCBHs (see Eq. 32 and Fig. 4), and is comparable to massive star clusters observed in the local Universe (e.g. Baumgardt & Hilker 2018; Krumholz et al. 2019; Neumayer et al. 2020; Brown & Gnedin 2021). The formation of SNF nuclei thus implies that a large amount of dense gas is available for star formation and BH growth, as to be described below.

4.3 The turbulence-modified Bondi accretion of black holes

4.3.1 The role of turbulence

The turbulent environment in the SGC provides a mechanism for the gas to efficiently flow through the SNF nucleus, cross the radius

of influence (i.e. Bondi radius) of the central BH, and eventually feed the BH. In the outer volume of an SGC, the formation and condensation of sub-clouds makes collision among them infrequent and drag force from surrounding gas negligible, as shown in §3.3 of Paper-I. This leads to the ballistic motion of sub-clouds, so that angular-momentum mixing between them is weak. Towards the inner volume of an SGC, however, the collision frequency for a sub-cloud, ν_{coll} , increases. The number of collision expected for a sub-cloud per dynamical time is

$$N_{\text{coll}} = \nu_{\text{coll}} t_{\text{dyn}} \propto V_{\text{c}} t_{\text{dyn}} R_{\text{sc}}^2 \rho_{\text{g}} \propto \rho_{\text{g}}^{1/2} R_{\text{sc}}^2 \propto r^{-1} R_{\text{sc}}^2, \quad (58)$$

where R_{sc} is the typical size of a sub-cloud, and an isothermal profile is assumed for ρ_{g} of the SGC. Thus, N_{coll} increases rapidly with decreasing r , unless R_{sc} shrinks. Of particular significance is the case when a sub-cloud is close to or within the SNF nucleus, $r \lesssim R_{\text{nuc}}$, where the environmental density is $\rho_{\text{g}} \approx \rho_{\text{snf}}$. Such a density is comparable to that of sub-clouds themselves, including even the densest ones that host globular clusters (see e.g. fig. 6 of Paper-III). Thus, sub-clouds in the nucleus volume are expected to interact frequent. A similar estimate can be made for the drag force by the ambient gas on a sub-cloud. The force is stronger at smaller r and becomes significant at $r \lesssim R_{\text{nuc}}$.

The dissipation has important consequences for gas dynamics in inner SGC. First, the collision between sub-clouds can compress the gas. Owing to the more effective cooling at higher gas density, the compressed gas does not push back, but becomes denser and more massive than the original sub-clouds. This leads to further fragmentation due to the reduced Jeans mass. Second, the loss of orbital energy and angular momentum due to the collision causes the gas to sink to a smaller orbit. Statistically, for a large number of sub-clouds that experience repeated collision, the orbit sinking generates a net inflow of gas from the outer part of the SGC to the nucleus, and from nucleus to the central BH. This can be understood by considering an extreme case where two sub-clouds have the same mass, move on circular orbits with the same speed but opposite direction. The head-on collision between them totally cancels the angular momentum, and enables the gas to fall freely towards the central BH. Thus, the net effect of collision of sub-clouds on the gas dynamics is to cause (i) mergers of sub-clouds to form denser ones; (ii) the conversion of kinetic energy to radiation; (iii) the loss of angular momentum; and (iv) mass inflow through the SNF nucleus towards the central BH. These conclusions hold even if the SGC is metal-poor, since the shock compression can raise the gas temperature above 10^4 K, the regime of efficient cooling by hydrogen and helium atoms. A similar conclusion can be reached by considering the interaction between sub-clouds and the ambient gas. Sub-clouds can survive the disruption by the surrounding gas due to cooling in the turbulence mixing layer, if the sub-cloud density is significantly higher than that of the ambient gas (e.g. Gronke & Oh 2020). The interaction with the surrounding gas thus provides ram-pressure that compresses sub-clouds, and a friction force that drags sub-clouds inwards.

The essential role of turbulence can thus be summarized as ‘deferred mixing’. In the outer part of an SGC, turbulence seeds the formation of sub-clouds and amplifies their density, so that the collision between sub-clouds and the drag force on them are both negligible. This preserves the broad distribution of j , allowing sub-clouds with low j to approach the central region of a halo without being mixed with sub-clouds with high j . However, as the low- j sub-clouds approach the inner SGC, collision and mixing between them become important. The efficient dissipation of energy and mixing of j can then drive a gas inflow to feed the central black hole. In the absence of a strong disturbance, the inflow is expected to be steady, produc-

ing a quasi-equilibrium state in which the gas inflow balances gas consumption by star formation, BH growth and associated feedback. Such a mode of BH growth, modeled in Paper-I, is referred to as the ‘*continuous mode*’ in the following. However, in the presence of a strong disturbance, such as that associated with a strong merger, the collision of sub-clouds and the mixing of angular momentum can occur over a large volume of the SGC. This produces a temporary surge in the gas inflow that can overwhelm the outflow generated by the feedback from the existing BH and stars, leading to the formation of a massive SNF nucleus that can power a ‘burst’ of BH growth. In this case, the feedback may become ‘positive’, since the outflow generated by it provides an additional source to compress and mix the gas. This mode of BH growth will be referred to as the ‘*bursty mode*’ in the following. The effect of the deferred mixing of sub-clouds on the distribution of j is schematically illustrated in Fig. 5a. A sharply peaked, centralized distribution forms via efficient mixing in the inner region of an SGC, on top of a broad distribution expected from the lack of mixing in the outer SGC. Note that the net angular momentum of the SGC is not necessarily different from that of a dynamically-cold disk. Instead, it is the turbulence that continuously redistributes j and generates low- j sub-clouds to feed the central region.

4.3.2 The turbulence-modified Bondi accretion rate

The SNF nuclei formed through the rapid collapse of sub-clouds into halo centers are expected to be dynamically hot and turbulent. This motivates the use of modified Bondi accretion (Hoyle & Lyttleton 1939; Bondi & Hoyle 1944; Bondi 1952) to model the amount of gas to be accreted into the zone of influence of the central BH. The isotropic condition of the Bondi accretion may be met here due to the dynamical hotness, but the velocity dispersion of gas elements is now dominated by the turbulent motion rather than the thermal motion. Modified by the turbulence, the Bondi accretion rate becomes:

$$\dot{M}_{\text{B}} = 4\pi R_{\text{B}}^2 \rho_{\text{g}}(R_{\text{B}}) \sigma_{\text{eff}}(R_{\text{B}}), \quad (59)$$

where σ_{eff} is the effective velocity dispersion of the gas defined by $\sigma_{\text{eff}}^2 = (c_{\text{s}}^2 + v_{\text{t}}^2)^{1/2}$, accounting for both thermal motion (quantified by the sound speed, c_{s}) and turbulent motion (quantified by turbulent velocity dispersion v_{t}). The Bondi radius, R_{B} , is the radius of the zone of influence of the BH, within which the gravity of the BH dominates the total gravity:

$$R_{\text{B}} = \frac{GM_{\text{BH}}}{\sigma_{\text{eff}}^2} = [0.430 \text{ pc}] M_{\text{BH},4} \sigma_{\text{eff},10}^{-2}, \quad (60)$$

where $\sigma_{\text{eff}} = 10 \text{ km s}^{-1} \sigma_{\text{eff},10}$.

In the supersonic, dispersion-dominated regime that is expected for an SNF nucleus,

$$V_{\text{c}} \approx \sigma_{\text{eff}} \approx v_{\text{t}} \gg c_{\text{s}}. \quad (61)$$

Using the general density profile of the SGC (Eq. 43), we can express the circular velocity profile at $r \ll R_{\text{nuc}}$ as

$$V_{\text{c}} = \psi_{\text{nuc}} V_{\text{nuc}} \left(\frac{r}{R_{\text{nuc}}} \right)^q, \quad (62)$$

where $\psi_{\text{nuc}} \sim \mathcal{O}(1)$ is a normalization factor. In the default case of $q = 1/2$, $\psi_{\text{nuc}} \approx 1.28$ (Eq. 49). The density profile is given by

$$\rho_{\text{g}}(r) = \frac{(2q+1)\psi_{\text{nuc}}^2 \Omega_{\text{nuc}}^2}{4\pi G} \left(\frac{r}{R_{\text{nuc}}} \right)^{2q-2}. \quad (63)$$

The Bondi radius can then be obtained from Eq. (60):

$$R_B = R_{\text{nuc}} \left(\frac{M_{\text{BH}}}{\psi_{\text{nuc}}^2 M_{\text{nuc}}} \right)^{1/(2q+1)}. \quad (64)$$

We define the velocity and angular frequency at R_B as the Bondi velocity and the Bondi angular frequency, respectively:

$$V_B \equiv V_C(R_B) = \psi_{\text{nuc}} V_{\text{nuc}} \left(\frac{R_B}{R_{\text{nuc}}} \right)^q; \quad (65)$$

$$\begin{aligned} \Omega_B &\equiv \frac{V_B}{R_B} = \psi_{\text{nuc}} \Omega_{\text{nuc}} \left(\frac{R_B}{R_{\text{nuc}}} \right)^{q-1} \\ &= \psi_{\text{nuc}} \Omega_{\text{nuc}} \left(\frac{M_{\text{BH}}}{\psi_{\text{nuc}}^2 M_{\text{nuc}}} \right)^{\gamma_{\text{nuc}}}, \end{aligned} \quad (66)$$

where $\gamma_{\text{nuc}} = (q - 1)/(2q + 1)$ ranges from -1 ($q = 0$) to $1/2$ ($q \rightarrow +\infty$). Using Eq. (59) for \dot{M}_B , we obtain the Eddington ratio of the Bondi accretion rate:

$$\begin{aligned} \dot{m}_B &\equiv \frac{\dot{M}_B}{\dot{M}_{\text{Edd}}} = (2q + 1) t_{\text{Sal}} \Omega_B \\ &= (2q + 1) \psi_{\text{nuc}} t_{\text{Sal}} \Omega_{\text{nuc}} \left(\frac{M_{\text{BH}}}{\psi_{\text{nuc}}^2 M_{\text{nuc}}} \right)^{\gamma_{\text{nuc}}}, \end{aligned} \quad (67)$$

where $\dot{M}_{\text{Edd}} \equiv M_{\text{BH}}/t_{\text{Sal}}$ is the Eddington accretion rate¹, and $t_{\text{Sal}} \equiv \sigma_{\text{T}} c / (4\pi G m_{\text{p}}) \approx 450 \text{ Myr}$ is the Salpeter timescale. Since $\Omega_{\text{nuc}} \sim 1 \text{ Myr}^{-1}$, the factor $(2q + 1) \psi_{\text{nuc}} t_{\text{Sal}} \Omega_{\text{nuc}}$ in \dot{m}_B is of the order of 10^3 . For the default choice ($q = 1/2$), $\dot{m}_B \gtrsim 10^3$ for $M_{\text{BH}} \lesssim M_{\text{nuc}}$ (or $R_B \lesssim R_{\text{nuc}}$; see Eq. 64). This implies a super-Eddington accretion of the BH in the beginning of nuclear burst, provided the gas flowing across the Bondi radius can reach the BH (see §4.4).

In the default case ($q = 1/2$, $\gamma_{\text{nuc}} = -1/4$), R_B and \dot{m}_B (Eqs. 64 and 64) can be expressed neatly as functions of Σ_{nuc} and M_{BH} . Using Eq. (56), the dependence on Σ_{nuc} can then be transformed into a dependence on Σ_{SGC} or M_V . This gives the following two chains of equations:

$$\begin{aligned} R_B &= \psi_{\text{nuc}}^{-1} \left(\frac{M_{\text{BH}}}{\Sigma_{\text{nuc}}} \right)^{1/2} \\ &= \psi_{\text{nuc}}^{-1} \left(\frac{3}{m_{\text{g}}(1)} \frac{M_{\text{BH}}}{c_{\text{g}} \Sigma_{\text{SGC}}} \right)^{1/2} \\ &\approx 4.56 \text{ pc} \frac{M_{\text{BH},6}^{1/2}}{[\chi_{\text{g}}(1+z)_{10}]^{1/4} M_{\text{v},9.5}^{1/6}}; \\ \dot{m}_B &= 2\psi_{\text{nuc}}^{3/2} t_{\text{Sal}} G^{1/2} \frac{\Sigma_{\text{nuc}}^{3/4}}{M_{\text{BH}}^{1/4}} \\ &= 2\psi_{\text{nuc}}^{3/2} t_{\text{Sal}} G^{1/2} \left[\frac{m_{\text{g}}(1)}{3} \right]^{3/4} \frac{(c_{\text{g}} \Sigma_{\text{SGC}})^{3/4}}{M_{\text{BH}}^{1/4}} \\ &\approx 6010 [\chi_{\text{g}}(1+z)_{10}]^{3/8} \frac{M_{\text{v},9.5}^{1/4}}{M_{\text{BH},6}^{1/4}}, \end{aligned} \quad (68)$$

Each of the chains is composed of three equations that link quantities at four different radii: the Bondi radius (R_B), the radius of the SNF

¹ Since we do not impose a constant value on the radiative efficiency (see §4.5), the definition of the Eddington rate of BH accretion in this paper assumes a radiative efficiency of unity for simplicity. We will clarify alternative definitions when discussing studies using them.

nucleus (R_{nuc}), the radius of the SGC (R_{SGC}), and the virial radius of the halo (R_V). Note that R_B depends weakly on M_V and z due to the weak dependence of Σ_{nuc} on M_V and z (see the texts below Eq. 56). This implies that the Bondi radius in an SNF nucleus is mainly controlled by the BH mass. The power-law indices on Σ_{nuc} in both chains are less than 1. Combined with the nearly flat evolution of Σ_{nuc} during the assembly history of any halo (see Fig. B1c), this implies that both R_B and \dot{m}_{BH} evolve very slowly for given M_{BH} and $M_{V,z=0}$.

The inner ends of the green segments in Fig. 6a show the predicted Bondi radii of BHs with $M_{\text{BH}} = 10^6 M_{\odot}$ embedded in the SNF nuclei that are expected to form in an MW-size halo and its progenitors. The Bondi radii are a few pc, and as expected, evolve very little from $z = 9$ to $z = 0$. For reference, we show the Bondi properties for wide ranges of halo masses and redshifts in Appendix B and Fig. B2.

4.4 Gas structure inside the Bondi radius

Within the Bondi radius, the central BH dominates the total gravity. Gas clouds moving across this radius is tidally torn apart, forming a new gaseous structure that fuels the BH. As suggested by hydro simulations (Anglés-Alcázar et al. 2021; Hopkins et al. 2024a,b; Shi et al. 2024a), the gas dynamics in this region is highly asymmetric, eventually leading to the build-up of an accretion disk surrounding the BH. In the regime of the high accretion rate expected in an SNF nucleus, the accretion disk does not resemble an α -disk (Shakura & Sunyaev 1973) in the following aspects (see Hopkins et al. 2024c, thereafter H24, for the details):

- (i) The disk is turbulent, with a sonic Mach number $\mathcal{M}_s = v_t/c_s \gg 1$, a consequence of the rapid (super-Eddington) inflow. In contrast, turbulence in an α -disk is subsonic.
- (ii) Magnetic field can be strong, with an Alfvén velocity $v_A \equiv B/\sqrt{4\pi\rho_{\text{g}}} \gg c_s$. The strong magnetic field appears to be inevitable. Even if the seed field is initially small, it can be amplified by the turbulent dynamo and reach to equal partition with the gas kinetic energy ($v_A \sim v_t$, with the Alfvén Mach number $\mathcal{M}_A \equiv v_t/v_A \sim 1$), over a number of dynamical timescales, as suggested by MHD simulations (Martin-Alvarez et al. 2018; Hopkins et al. 2024a) and analytical studies (Pariev et al. 2003; Begelman & Pringle 2007; Begelman & Silk 2023).
- (iii) The disk is thickened by the large v_t and v_A , in contrast to an α -disk which is thin ($H_{\text{g}}/r \ll 1$) and supported by thermal and radiation pressure.
- (iv) The gas inflow rate in the disk towards the BH, \dot{M}_{g} , is large, which is a consequence of the large kinematic viscosity contributed by the large v_t^2 (Reynolds stress) and v_A^2 (Maxwell stress).
- (v) The disk is stable against fragmentation and star formation, as the Toomre Q parameter is expected to be large (see Eq. 83).

The transition of the gas from the less-structured turbulence at $r \gtrsim R_B$ to an ordered accretion disk at $r \lesssim R_B$ implies an additional level in the cosmic hierarchy. In particular, the properties of the accretion disk determine the possibility of star formation of the gas around the BH and eventually the growth rate of the central BH. We thus explicitly include the accretion disk in our modeling, with the outer boundary conditions set by the Bondi properties at R_B obtained in §4.3 and the inner boundary conditions set by the BH itself. Briefly, the boundary conditions include M_{BH} of the BH, \dot{M}_{g} , v_t/V_K and v_A/V_K , all evaluated at R_B , and some normalization factors. Here, $V_K \approx V_c$ is the Keplerian velocity, which approximates the circular velocity at $r \leq R_B$ as the gravity is dominated by the BH in this region. The output of this model is v_t/V_K , v_A/V_K and Σ_{g} , and other variables deducible from them, such as the surface gas density of

the disk (Σ_g), the disk height (H_g), and the Toomre Q parameter, all as functions of the BH-centric distance r within the domain of $r_s \ll r \leq R_B$. Here r_s is the Schwarzschild radius, defined as

$$R_s = \frac{2GM_{\text{BH}}}{c^2} = 10^{-4} \text{ pc } M_{\text{BH},9} = 20 \text{ au } M_{\text{BH},9} = 0.02 \text{ au } M_{\text{BH},6}, \quad (70)$$

at which relativistic effects become significant. The model we adopt is developed by H24, which was shown to be capable of reproducing results of their hyper-refinement zoom-in simulations carried out in a realistic cosmological context. We refer the readers to their paper for details.

The basic equation can be derived by considering the energy conservation that dissipation of the turbulence/magnetic energy within the disk is driven by the release of the gravitational energy of the inflow gas. Equating the dissipation rate per unit area, $dE_{\text{t+A}}/(dA dt)$, and the gravitational power per unit area, $dE_{\text{grav}}/(dA dt)$, we obtain

$$\frac{3}{2\psi_t^2} \Sigma_g \Omega_K v_t^2 = \frac{3}{4\pi} \Omega_K^2 \dot{M}_g. \quad (71)$$

Here, $\Omega_K = V_K/r$ is the angular frequency of a circular orbit, and \dot{M}_g is the net inflow (inflow minus outflow) rate of the gas. As we will see below (Eq. 83), the accretion disk does not have significant star formation, and thus the Bondi accretion rate is preserved across the accretion disk, i.e. $\dot{M}_g = \dot{M}_B = \dot{M}_{\text{BH}}$. The condition $\psi_t^2 \sim \mathcal{O}(1)$ fuses the normalization factors of both sides in the equation. The total specific energy to dissipate is about $(1/2)(v_t^2 + v_A^2) \sim v_t^2$, where the approximation holds for a magnetic field close to the equal-partition level. The dissipation timescale is $2\psi_t^2/(3\Omega_K)$. More formally, this equation has to be derived by the conservation of angular momentum (see Pariev et al. 2003, for details). An equivalent formalism is $\dot{M}_g = 3\pi\Sigma v_{\text{visc}}$, where the viscosity is $v_{\text{visc}} = 2v_t^2/(3\psi_t^2\Omega_K)$. The inflow velocity, $|v_r|$, can be deduced by its definition,

$$|v_r| \equiv \frac{\dot{M}_g}{2\pi r \Sigma_g} = \left(\frac{v_t}{\psi_t V_K} \right)^2 V_K. \quad (72)$$

To form a closed set for Eq. (71), we need an equation to describe the turbulence structure, $v_t(r)$. Here we seek for similarity solutions and assume that the radial dependence of the turbulent velocity can be parameterized as $v_t \propto r^{-\zeta_t}$. We further assume that the gas falls freely at R_B , i.e. $|v_r(R_B)| = V_K$, which is used to fix the normalization. With these assumptions, the equation for the turbulence structure is

$$\frac{v_t}{V_K} = \psi_t y_r^{1/2-\zeta_t}, \quad (73)$$

where $y_r \equiv r/R_B$. Substituting Eq. (73) into Eq. (71), we obtain

$$\Sigma_g = \frac{\dot{M}_g}{2\pi\sqrt{GM_{\text{BH}}R_B}} y_r^{2\zeta_t-3/2}. \quad (74)$$

Interestingly, Σ_g does not depend on ψ_t .

The vertical structure of the disk, as suggested by H24, is contributed by both turbulence and magnetic field, with the latter being even more dominant at smaller radius. The vertical structure of the disk is thus best parameterized through the magnetic field, as $H_g/r \sim v_A/V_K$. Based on their simulations, Hopkins et al. (2024a,b) found that the magnetic field within R_B can be described by the flux-frozen/advection scenario extrapolated from the outer boundary conditions set at R_B . They thus proposed an effective equation of state (eEOS) for the magnetic field, as $B^2 \propto \rho_g^{\gamma_B}$, or $v_A \propto \rho_g^{(\gamma_B-1)/2}$, where $\rho_g \sim \Sigma_g/(2H_g) \sim \Sigma_g\Omega_K/(2v_A)$ is the density at the mid-plane

of the accretion disk. Reorganizing v_A to one side of the equation, and using Eqs. (71) and (73) for Σ_g , we obtain

$$\frac{v_A}{V_K} = \psi_A y_r^{1/2-\zeta_A}, \quad (75)$$

where $\zeta_A = (3-2\zeta_t)(\gamma_B-1)/(\gamma_B+1)$, and ψ_A is a normalization factor. With this equation, the height of the disk is $H_g \sim \psi_A R_B y_r^{3/2-\zeta_A}$.

Eqs. (71), (73) and (75) form a closed set describing the density, turbulence and magnetic field structure of the accretion disk. The set is fixed once the structure index ζ_t and ζ_A (or γ_B) are given/calibrated, and the outer boundary R_B and the boundary conditions on M_{BH} , \dot{M}_B , ψ_t and ψ_A are obtained from the modeling described earlier. Other by-products can be derived using these equations, such as the Toomre Q parameter of the disk:

$$Q \equiv \frac{\sigma_{\text{eff}\kappa}}{\pi G \Sigma_g} \approx 2\sqrt{2}\psi_t \sqrt{1 + \mathcal{M}_A^{-2} \frac{M_{\text{BH}}\Omega_B}{\dot{M}_g}} y_r^{-3\zeta_t}, \quad (76)$$

where we have used $\kappa = \sqrt{2}\Omega_K$, $\Omega_B = \Omega_K(R_B)$, and $\sigma_{\text{eff}}^2 = v_t^2 + v_A^2$, with the sound speed neglected.

A special case of this model (the default model of H24), which we adopt in this paper, is given by a flux-frozen magnetic field ($\gamma_B = 4/3$) and a trans-Alfvénic, supersonic turbulence ($\mathcal{M}_A \sim 1$, $\mathcal{M}_s \gg 1$). These imply that $\zeta \equiv \zeta_t = \zeta_A = 1/3$, $\psi \equiv \psi_t \sim \psi_A$, and thus

$$\frac{H_g}{r} \sim \frac{v_A}{v_K} \sim \frac{v_t}{v_K} = \psi y_r^{1/6}; \quad (77)$$

$$\Sigma_g = \frac{\dot{M}_g}{2\pi\sqrt{GM_{\text{BH}}R_B} y_r^{5/6}}; \quad (78)$$

$$\rho_g \sim \frac{\dot{M}_g}{4\pi\psi\Omega_B R_B^3 y_r^2}; \quad (79)$$

$$Q \sim \frac{4\psi M_{\text{BH}}\Omega_B}{\dot{M}_g y_r}. \quad (80)$$

Taking R_B from Eq. (68), we obtain

$$\Sigma_g = 24.7 M_\odot \text{pc}^{-2} \frac{M_{\text{BH},6}^{1/2} \dot{m}_{\text{B},10}}{R_{\text{B},4.56}^{1/2} y_r^{5/6}} \quad (81)$$

$$= 24.7 M_\odot \text{pc}^{-2} [\chi_g(1+z)_{10}]^{1/8} \frac{M_{\text{v},9.5}^{1/12} M_{\text{BH},6}^{1/4} \dot{m}_{\text{B},10}}{y_r^{5/6}};$$

$$\rho_g \sim 2.70 M_\odot \text{pc}^{-3} \psi^{-1} \frac{M_{\text{BH},6}^{1/2} \dot{m}_{\text{B},10}}{R_{\text{B},4.56}^{3/2} y_r^2} \quad (82)$$

$$= 2.70 M_\odot \text{pc}^{-3} \psi^{-1} [\chi_g(1+z)_{10}]^{3/8} \frac{M_{\text{v},9.5}^{1/4} \dot{m}_{\text{B},10}}{M_{\text{BH},6}^{1/4} y_r^2};$$

$$Q \sim 1240\psi \frac{M_{\text{BH},6}^{1/2}}{R_{\text{B},4.56}^{3/2} \dot{m}_{\text{B},10} y_r} \quad (83)$$

$$= 1240\psi [\chi_g(1+z)_{10}]^{3/8} \frac{M_{\text{v},9.5}^{1/4}}{M_{\text{BH},6}^{1/4} \dot{m}_{\text{B},10} y_r},$$

where $\dot{m}_{\text{B},10} \equiv \dot{m}/10$ is the Eddington ratio of the Bondi accretion rate assuming a radiative efficiency of $\chi_{\text{AGN}} = 0.1$ (Crain et al. 2015; Rosas-Guevara et al. 2015). The properties of the accretion disk depend weakly on M_v and M_{BH} , but strongly on \dot{m}_B and y_r . $Q \gg 1$ is implied by Eq. (83) at any $y_r \leq 1$, consistent with the assumption that the disk is stable against fragmentation (see the texts below Eq. 71).

This is true even when we consider the extremely high accretion rate of $\dot{m}_B \sim 6000$ at the beginning of the nuclear burst (Eq. 69).

Fig. 5e schematically shows the structures within the Bondi radius – the last level in the cosmic hierarchy. The accretion disk is turbulent, magnetized and thickened. Star formation is suppressed in the inner volume of this regime, considering the r^{-1} dependence of the Toomre Q parameter (Eq. 83). Stars might be present in the outer volume, which is not properly captured by the approximation of the accretion-disk model. These stars might also redistribute their energy and angular momentum via dynamical processes, and contribute to the BH growth. Such contribution is referred to as the ‘dry’ channel, in complement to the ‘wet’ channel of gas accretion. Each orange segment in Fig. 6a shows the mid-plane density of the accretion disk formed during the nuclear burst in an MW-size halo or its progenitors, assuming a central BH with $M_{BH} = 10^6 M_\odot$ and a maximal accretion rate given by Eq. (69) at the beginning of the burst. The orange shade covers the range of ρ_g as \dot{m}_{BH} varies from the Eddington level ($\dot{m}_{BH} = 10$) to the maximum. The profile of ρ_g varies very little from $z = 0$ to $z = 10$, due to the weak evolution of the Bondi radius and Bondi accretion rate (see the texts below Eqs. 68 and 69). For reference, the card in Fig. 6e summarizes the equations for the accretion disk.

The role of turbulence in BH growth is again highlighted here. It magnifies the magnetic field, suppresses star formation, and preserves the rapid inflow at R_B through the accretion disk to feed the BH. As outlined in the beginning of §4.2, such a rapid process is expected to be transient, associated with bursty star formation, BH accretion and feedback. The evolution of the SNF nucleus in the dispersion stage is thus a process expected to be largely asynchronous with the evolution of the host galaxy/halo. This suggests that a refinement engine working at much finer spatial scales and temporal steps than those for modeling the host galaxy/halo is needed to follow the evolution of the hierarchy below the SNF nuclei, as to be described in the following.

4.5 The bursty nature of black-hole growth and star formation in the nuclei

Here we build a model for the evolution of the SNF nucleus in the dispersion stage. Nuclear burst, including the bursty star formation and BH accretion, and their feedback are modeled together. A number of depletion timescales in the nucleus are analytically derived and compared with each other to gain insights into the mechanisms driving the evolution. The equations governing the evolution can be solved numerically, and will be incorporated as a refinement engine to complete the whole modeling procedure (see §5.1 and Fig. 8 for an outline).

4.5.1 Basic considerations

Simplifications are necessary for the model to be applicable in a cosmological volume, while the essence of the nuclear burst must be retained. Here we model the distributions of the gas and stars within the SNF nucleus using spherically averaged, smoothed profiles. This takes account of the isotropy expected, at least statistically, in a dynamically hot nucleus. Sub-structures, such as gas and stellar clumps, are missed in such a treatment. Thus, star formation, gas inflow and interaction between feedback and gas in the model are treated as effective processes representing the average over diffuse and clumpy structures.

The basic equation is the continuity equation of gas density

$\rho_g(r, t)$:

$$\frac{\partial \rho_g}{\partial t} = -\frac{\partial \rho_*}{\partial t} - \frac{(\dot{e}_{AGN} + \dot{e}_{SN} + \dot{e}_{SE} + \dot{e}_{ext}) \rho_g}{v_{out}^2} - \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \rho_g v_r), \quad (84)$$

where v_r is the radial velocity of inflow. This equation describes the reduction of the gas density at a given BH-centric distance r due to different sinks, each quantified by one term in the right-hand side. Here, $\rho_*(r, t)$ is the stellar density, and the term $-\partial \rho_*/\partial t$ describes the consumption of the gas by star formation. The terms, \dot{e}_{AGN} , \dot{e}_{SN} , \dot{e}_{SE} , are the specific (i.e. per unit mass) energy injection rates by the feedback from the BH, SN explosions and stellar evolution within the nucleus, respectively, while \dot{e}_{ext} is from the star formation outside the nucleus. The quantity v_{out} denotes the velocity of the outflow driven by all the feedback effects. The total baryon density will be denoted as $\rho = \rho_g + \rho_*$. The time domain of this equation is $t \geq t_{burst}$, with t_{burst} being the starting epoch of the dispersion stage. The radial domain of the equation is $R_{in} \leq r \leq R_{nuc}$, with the inner boundary at $R_{in} \ll R_B$ in the beginning so that the time evolution of the Bondi radius can be self-consistently determined by its definition (Eq. 60) and the structure equation at $r \leq R_B$ can be determined by the model of the accretion disk (§4.4).

To estimate of the relative importance of each term in Eq. (84), we consider the depletion timescale, $t_{dep,x} = M_{nuc}/\dot{M}_x$, for a process x that changes the gas mass at a rate of \dot{M}_x . The last sink term in Eq. (84) corresponds to a continuous accretion flow through the nucleus towards the BH. The corresponding mass inflow rate can be estimated using the Bondi accretion rate, \dot{M}_B (Eq. 67), and the depletion timescale is:

$$t_{dep,bhacc} = \frac{M_{nuc}}{\dot{M}_B} = \frac{M_{nuc}}{(2q+1)\Omega_B M_{BH}}. \quad (85)$$

The first sink term in Eq. (84) describes the gas consumption by star formation. The corresponding depletion timescale may be approximated by the free-fall time of the gas within the nucleus (Eq. 54),

$$t_{dep,SF} = t_{ff,nuc} = \frac{\pi}{2\sqrt{2}\Omega_{nuc}}. \quad (86)$$

The ratio between the two depletion timescales is

$$\frac{t_{dep,bhacc}}{t_{dep,SF}} = \frac{2\sqrt{2}\psi_{nuc}^{2\gamma_{nuc}-1}}{\pi(2q+1)} \left(\frac{M_{nuc}}{M_{BH}} \right)^{\gamma_{nuc}+1}. \quad (87)$$

The prefactor is of the order of unity, while the mass-dependence is $(M_{nuc}/M_{BH})^{3/4}$ in the default case ($q = 1/2$). The timescale $t_{dep,SF} \ll t_{dep,bhacc}$ if $M_{BH} \ll M_{nuc}$, and $t_{dep,SF} \gg t_{dep,bhacc}$ if $M_{BH} \gg M_{nuc}$. This suggests a competitive scenario between BH accretion and star formation in the SNF nucleus, with the dominator determined by M_{nuc}/M_{BH} .

The second sink term in Eq. (84) groups the effects of different feedback channels of gas depletion. The energy injection by the SN feedback, \dot{e}_{SN} , is expected to be negligible before SN explosions from dying massive stars ($t - t_{burst} \lesssim t_{snf}$; see the texts below Eq. 41), but to be sufficiently strong afterwards to totally disperse the gas from the nucleus. SNe thus place a hard limit on the duration of the nuclear burst. The depletion timescale by the SN feedback is thus $t_{dep,SN} = t_{snf}$, comparable to that by the star formation ($t_{dep,SF}$). The energy injection by stellar evolution, \dot{e}_{SE} , and by external star formation, \dot{e}_{ext} , are expected to be negligible owing to the high density of the nucleus, as discussed in §4.2. Thus, during a burst

of an SNF nucleus, AGN feedback is the only possible channel of feedback that is effective to deplete gas.

The exact effects of AGN feedback on the nuclear gas, however, are difficult to quantify from first principles. The difficulty comes from uncertainties in, e.g. the ‘structure’ of the BH (e.g. spin; see [Huško et al. 2022](#)), the state and emissivity of the gas in the accretion disk ([Yuan & Narayan 2014](#); [Koudmani et al. 2024](#); [Hopkins et al. 2024c](#)) and in the circum-BH medium ([Dutta et al. 2024](#)), the potential launch of ultra-fast outflows ([Wagner et al. 2013](#); [Gaspari & Sądowski 2017](#)), the coupling of the feedback energy (in radiation and outflow) with the nuclear gas ([Su et al. 2023, 2025](#)), and the interplay among them. To simplify the problem, we do not attempt to differentiate channels of the AGN feedback (e.g. jet, wind, radiation and cosmic ray; see [Shi et al. 2024b](#)). Instead, we parameterize the total mass loss rate of the SNF nucleus produced by the AGN feedback as

$$\dot{M}_{\text{AGN}} = \chi_{\text{AGN}} L_{\text{Edd}} / v_{\text{out}}^2, \quad (88)$$

where $L_{\text{Edd}} \equiv M_{\text{BH}} c^2 / t_{\text{Sal}}$ is the Eddington luminosity; χ_{AGN} is the Eddington ratio of feedback energy, taking into account the net effect from both the energy conversion efficiency of the BH accretion and the coupling efficiency of the feedback energy to the nuclear gas. As suggested by e.g. [Inayoshi et al. \(2016\)](#), a thickened accretion disk in the super-Eddington regime is radiatively inefficient due to ‘photon trapping’, and is able to radiate at the Eddington level. Given the high density in the SNF regime, most of the emitted radiation is expected to be coupled with the nuclear gas. This implies the AGN feedback is inefficient in radiating but efficient in coupling, very different from the sub-grid recipes commonly implemented in unresolved hydrodynamic simulations (e.g. [Schaye et al. 2015](#); [Weinberger et al. 2017](#); [Davé et al. 2019](#)). We thus adopt $\chi_{\text{AGN}} = \chi_{\text{AGN},\infty} = 1$ as our default value at the high accretion rate (see below for the treatment of low accretion rate). The default outflow velocity is $v_{\text{out}} = 10^3 \text{ km s}^{-1}$, as suggested by observations (e.g. [Sturm et al. 2011](#); [Greene et al. 2012](#); [Perna et al. 2017, 2019](#); [D’Eugenio et al. 2024](#)) and feedback models (e.g. [Costa et al. 2014](#); [Davé et al. 2019](#)). With the adopted parameterization, the depletion timescale for the AGN feedback is

$$t_{\text{dep,AGN}} = \frac{M_{\text{nuc}}}{\dot{M}_{\text{AGN}}} = \frac{t_{\text{Sal}}}{\chi_{\text{AGN}}} \frac{M_{\text{nuc}}}{M_{\text{BH}}} \left(\frac{v_{\text{out}}}{c} \right)^2. \quad (89)$$

The ratio of the depletion timescale between the BH accretion and the AGN feedback is then

$$\frac{t_{\text{dep,bhacc}}}{t_{\text{dep,AGN}}} = \frac{\psi_{\text{nuc}}^2 \gamma_{\text{nuc}}^{-1} \chi_{\text{AGN}}}{(2q+1) t_{\text{Sal}} \Omega_{\text{nuc}}} \left(\frac{c}{v_{\text{out}}} \right)^2 \left(\frac{M_{\text{nuc}}}{M_{\text{BH}}} \right)^{\gamma_{\text{nuc}}}, \quad (90)$$

where $(c/v_{\text{out}})^2 \approx 300^2$ and $t_{\text{Sal}} \Omega_{\text{nuc}} \approx 450 \times 1.8 \approx 800$. In the default case ($q = 1/2$, $\gamma_{\text{BH}} = -1/4$), the factor $(M_{\text{nuc}}/M_{\text{BH}})^{\gamma_{\text{nuc}}} \geq 0.01$ for any $M_{\text{BH}}/M_{\text{nuc}} \geq 10^{-8}$. Thus, $t_{\text{dep,bhacc}} \gg t_{\text{dep,AGN}}$ almost always holds.

The above estimates yield a number of conclusions: (i) the depletion of gas in an SNF nucleus is dominated by the AGN feedback; (ii) if the AGN feedback does not disperse all the gas until $t - t_{\text{burst}} = t_{\text{snf}}$, SN feedback kicks in and clears the remaining gas; (iii) a short time window, $t - t_{\text{burst}} \in [0, t_{\text{snf}}]$, is left for intensive BH accretion and star formation, which leads to bursty growths in mass for both the BH and in the stellar component. The temporary absence of SN feedback in the time window thus gives the BH a chance to grow differently from that expected in the continuous mode where SNe constantly regulate the gas content around the BH. Such a bursty mode may be missed in numerical simulations that are incapable of resolving dense structures at $n_{\text{g}} \gg n_{\text{snf}}$. The gas medium around BHs in these simulations may be numerically over-smoothed, so that the SN feedback

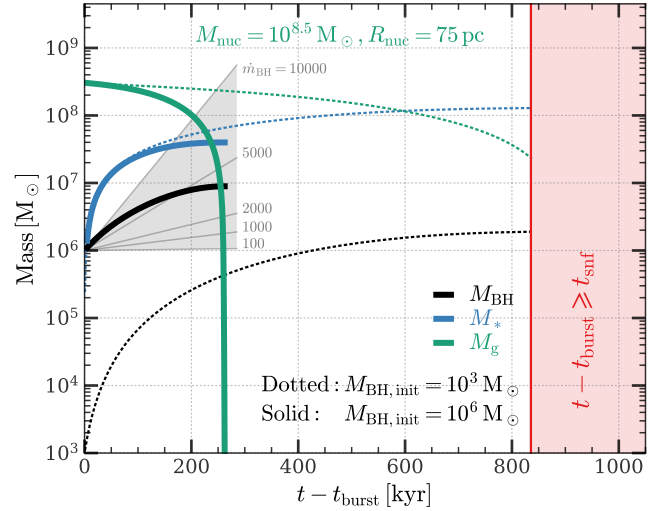


Figure 7. Evolution of masses in nuclear bursts predicted by the refinement engine. Here we show two events of nuclear bursts for example, both with a BH placed in the center of the SNF nucleus with $M_{\text{nuc}} = 10^{8.5} M_{\odot}$ and $R_{\text{nuc}} = 75 \text{ pc}$, typical for MW-size galaxies and their progenitors (see e.g. Fig. B1). The initial BH mass is $M_{\text{BH}} = 10^3 M_{\odot}$ (dotted) and $10^6 M_{\odot}$ (solid), respectively, in the two bursts. In each burst, **black**, **blue** and **green** curves show the masses of the BH, stars and gas, respectively, as a function of elapsed time since the beginning of the burst. **Red** shaded region indicates the truncation of the burst by SN feedback, which sets a time window with a duration of t_{snf} for the BH growth and star formation. **Grey** lines in grey shaded region represent the evolution tracks of BH expected with constant Eddington ratios, $\dot{m}_{\text{BH}} \equiv \dot{M}_{\text{BH}}/\dot{M}_{\text{Edd}}$. The numerical solver will be incorporated into the model as a refinement engine to obtain the evolution of masses in each nuclear burst. See §4.5 for details.

is always effective to regulate the BH growth (see e.g. [Bower et al. 2017](#); [Habouzit et al. 2017, 2021](#); [Li et al. 2024](#)). In the following, we present a detailed numerical treatment based on these conclusions.

4.5.2 A refined treatment

To solve $\rho_{\text{g}}(r, t)$ and $\rho_{\text{*}}(r, t)$ for an SNF nucleus during the dispersion stage, we need to explicitly formulate each sink term in Eq. (84). The term of star formation is parameterized as

$$\frac{\partial \rho_{\text{*}}}{\partial t} = \frac{\alpha_{\text{SF}} \rho_{\text{g}}}{f_{\text{Q}} t_{\text{ff}}}, \quad (91)$$

where $\alpha_{\text{SF}} \sim 1$ is a free parameter, $t_{\text{ff}} = \sqrt{3\pi/(32G\rho)}$ is the free-fall timescale of gas. The suppression factor,

$$f_{\text{Q}} = \exp\left(\frac{Q - Q_{\text{B}}}{\delta_{\text{Q}}}\right), \quad (92)$$

takes account of the suppression of star formation by the large Toomre Q parameter in the accretion disk, which allows a rapid but continuous decrease of the star formation towards the center of the accretion disk; $\delta_{\text{Q}} \sim 1$ controls the sharpness of the decrease; Q is obtained from Eq. (80) at $r \leq R_{\text{B}}$, and $Q_{\text{B}} \equiv Q(R_{\text{B}})$; we set $Q = Q_{\text{B}}$ for $r > R_{\text{B}}$, meaning no suppression outside the Bondi radius.

The inflow rate at the Bondi radius is based on the turbulence-modified Bondi accretion (Eq. 59), with a free parameter $\alpha_{\text{BH}} \sim 1$ characterizing the normalization:

$$\dot{M}_{\text{BH}} = \alpha_{\text{BH}} \dot{M}_{\text{B}}. \quad (93)$$

The inflow term $\partial(r^2 \rho_{\text{g}} v_r)/\partial r / r^2$ in Eq. (84) depends on conditions

at the outer boundary, R_{nuc} , and the dissipation processes within this boundary. Here we model it by a stable solution, $\dot{M}_{\text{g}}(r) \equiv \dot{M}_{\text{BH}}$, in which the inflow rate $\dot{M}_{\text{g}}(r)$ is constant over the entire domain of $r \leq R_{\text{nuc}}$. This solution implies that the inflow term in Eq. (84) vanishes and the density profile, without being affected by star formation or feedback, does not change with time. This solution relies on the findings in §4.5.1 that the feedback term dominates the gas depletion at $R_{\text{B}} < r \leq R_{\text{nuc}}$ so that an arbitrary choice on the inflow term has negligible effect, and that star formation is suppressed at $r \leq R_{\text{B}}$ so that inflow rate is preserved throughout the zone of influence of the BH (see §4.4).

The specific energy injection rate by SNe is modeled by a delayed flare, as

$$\dot{\epsilon}_{\text{SN}}(t) = \begin{cases} 0, & \text{if } t - t_{\text{burst}} \leq t_{\text{snf}}; \\ \infty, & \text{otherwise,} \end{cases} \quad (94)$$

so that a hard limit of t_{snf} is set for a complete ejection of the gas in the SNF nucleus and the end of the dispersion stage. Other forms of stellar feedback, such as those represented by $\dot{\epsilon}_{\text{SE}}$ and $\dot{\epsilon}_{\text{ext}}$, are expected to be negligible and are ignored.

The specific energy injection rate by AGN is determined implicitly according to the parameterization of the total feedback energy (Eq. 88), as

$$\chi_{\text{AGN}} L_{\text{Edd}} = \int_{R_{\text{in}}}^{R_{\text{nuc}}} \dot{\epsilon}_{\text{AGN}} \rho_{\text{g}} 4\pi r^2 dr, \quad (95)$$

corresponding to a specific mass loss rate of $\dot{\epsilon}_{\text{AGN}}/v_{\text{out}}^2$. The Eddington ratio of the feedback energy, χ_{AGN} , is parameterized as a function of the Eddington ratio of the accretion rate:

$$\chi_{\text{AGN}} = \chi_{\text{AGN},\infty} \frac{\alpha_{\chi} \dot{m}_{\text{B}}}{1 + \alpha_{\chi} \dot{m}_{\text{B}}}, \quad (96)$$

so that χ_{AGN} is saturated at $\chi_{\text{AGN},\infty}$ for $\dot{m}_{\text{B}} \gg 1$, and $\alpha_{\chi} \sim 0.1$ in the sub-Eddington regime where the accretion disk is expected to become a normal α -disk (Shakura & Sunyaev 1973) that is often used as a sub-grid model in cosmological hydrodynamic simulations. The radial dependence of $\dot{\epsilon}_{\text{AGN}}$ is determined by the details of how the outflow mass and energy are coupled to the nuclear gas. Since these are largely unknown, we parameterize $\dot{\epsilon}_{\text{AGN}}$ as

$$\dot{\epsilon}_{\text{AGN}} \propto \rho_{\text{g}}^{-\beta_{\text{out}}}, \quad (97)$$

which accounts for the expectation that both coupling and cooling (dissipation) should depend on the gas density. We adopt $\beta_{\text{out}} = 0$ as the default, which implies a radially invariant specific mass loss rate caused by the AGN feedback. The calibration using hydrodynamic simulations, and the quantification for the effects of other relevant state variables, such as gas clumpiness and metallicity, are worthy of being explored further. Note that the outflow launched from the SNF nucleus can be very strong due to the Eddington-level luminosity of the BH. For $M_{\text{nuc}} \sim 10^8 M_{\odot}$ and $t_{\text{dep}} \lesssim 1$ Myr, we obtain $\dot{M}_{\text{w}} \gtrsim 100 M_{\odot} \text{yr}^{-1}$. When propagating outwards, the outflow can continuously couple with the gas in the SGC, and potentially launch galaxy-wide outflows.

Fig. 7 shows two examples of nuclear bursts. In each, a BH with a given initial mass is placed at the center of the SNF nucleus that is typical for MW-size galaxies and their progenitors. The evolution of mass components (BH, stars and gas) are shown as functions of elapsed time since the beginning of the burst. In both cases, the reduction of gas mass is substantial, much larger than the growth of BH and stellar masses. The main source of gas depletion is the AGN feedback. In the case of small initial BH mass, $M_{\text{BH,init}} = 10^3 M_{\odot}$,

about 1/10 of the gas remains in the nucleus at ≈ 800 kyr before SNe kick in to clear the gas. The central BH grows by more than three orders of magnitude in mass, from $10^3 M_{\odot}$ to more than $10^6 M_{\odot}$. A nuclear star cluster with $M_{*} > 10^8 M_{\odot}$ also forms in the burst, indicating a strong competition with the BH in consuming the gas. In the case with $M_{\text{BH,init}} = 10^6 M_{\odot}$, the feedback efficiency (energy output per unit accreted mass) is much higher, and the gas depletion terminates the burst at ≈ 200 kyr. Star formation here is less competitive than in the former case, and the growth of both the BH and stellar mass are about $10^7 M_{\odot}$. The final M_{BH} is higher than in the former case, but the final stellar mass is lower. The competition between the BH growth and star formation thus depends strongly on $M_{\text{nuc}}/M_{\text{BH}}$, consistent with the expectation based on the depletion timescales (Eq. 87). For comparison, grey lines in Fig. 7 show the evolution of the BH mass expected from a constant Eddington ratio (\dot{m}_{BH}) between 10^2 and 10^4 . The average growth rates of the two cases during the entire bursts fall in the range of $\dot{m}_{\text{BH}} = 10^3$ to 10^4 , suggesting that the bursty growth of the BH is super-Eddington over a wide range in the parameter space. Interestingly, observations indeed indicate that a small fraction of AGNs are experiencing super-Eddington accretion, either based on measurements of the BH mass and accretion rate (Liu et al. 2021), or based on the weak/variable X-ray emission expected from the obscuration of thick disk and/or clumpy wind surrounding a BH with super-Eddington accretion (Wang et al. 2022; Zhang et al. 2025b).

The diversities in the mass ratio, $M_{\text{nuc}}/M_{\text{BH}}$, in the efficiency of the AGN feedback and in the BH growth-star formation competition are thus expected to produce diversities in the growth path among different galaxy/BH systems. This can be explored further to explain the observed diversity of high- z galaxies/AGNs (e.g. Mezcua et al. 2024; Onoue et al. 2024; Li et al. 2025a), and our refined treatment for nuclear bursts provides a basis for such explorations within the current cosmological framework, as we will show in §5.

4.5.3 Comparison with other growth channels of black holes

To see that the nuclear burst is indeed an efficient channel of BH growth, here we compare it with other viable channels.

The continuous mode of BH growth, as derived in Paper-I using the loss-cone argument (Shapiro & Teukolsky 1986), predicts an accretion rate of $\dot{M}_{\text{BH}} \propto (M_{\text{BH}}/M_{\text{SGC}}) \dot{M}_{\text{g,cool}}$. Thus, the Eddington ratio of BH accretion rate is

$$\begin{aligned} \dot{m}_{\text{BH}} &\equiv \frac{\dot{M}_{\text{BH}}}{\dot{M}_{\text{Edd}}} = \alpha_{\text{cap}} t_{\text{sal}} \frac{F_{\text{cool}} f_{\text{b}}}{f_{\text{gas}}} \frac{\dot{M}_{\text{v}}}{M_{\text{v}}} \\ &= 3.67 \alpha_{\text{cap},2.5} \frac{F_{\text{cool}} f_{\text{b}}}{f_{\text{gas}}} M_{\text{v},9.5}^{0.14} (1+z)_{10}^{2.5}, \end{aligned} \quad (98)$$

where we have adopted Eq. (8) for an approximation of the halo assembly rate, \dot{M}_{v} ; F_{cool} is fraction of gas associated with the halo assembly that can cool and flow into the SGC; $\alpha_{\text{cap}} \equiv 2.5 \alpha_{\text{cap},2.5}$ is a free parameter, adopted to be 2.5 by default. Most halos at $z \approx 10$ have mass $M_{\text{v}} < 10^{12} M_{\odot}$ (see Fig. 1), implying effective cooling ($t_{\text{cool}} \lesssim t_{\text{v}}$; see fig. 2 of Paper-I) that leads to $F_{\text{cool}} \approx 1$ and $f_{\text{gas}} \approx f_{\text{b}}$. The dependence of \dot{m}_{BH} on halo mass is quite weak, while the factor $(1+z)^{2.5}$ implies a rapid decline of \dot{m}_{BH} with cosmic time. The expected value of \dot{m}_{BH} is $\lesssim 5$ at $z \approx 10$, marginally touching the Eddington rate, and falls rapidly to $\lesssim 0.1$ at $z \lesssim 3$. This implies a much less efficient accretion for BHs in the continuous mode than in the bursty mode. The relation between the BH growth and the halo assembly implied by Eq. (98) is $\dot{M}_{\text{BH}}/M_{\text{BH}} \propto \dot{M}_{\text{v}}/M_{\text{v}}$ and that between the BH growth and the star formation is $\dot{M}_{\text{BH}}/M_{\text{BH}} \propto \dot{M}_{*}/M_{*}$ if star formation can catch up with the BH accretion in the early

Universe. Given the relation between the BH mass and the galaxy mass, $M_{\text{BH}} \propto M_*^\gamma$ with the observed range of $\gamma \gtrsim 1$ in the local Universe (Kormendy & Ho 2013; Greene et al. 2020; Graham & Sahu 2023, e.g.), the continuous mode alone is not sufficient to drive the BH seed onto the observed relation, unless the seed is produced above the relation. In contrast, the bursty mode may be able to drive the BH seed onto or above the relation, as we will discuss in Appendix C.

In the case where turbulence is negligible at the scale of the SGC, rotationally supported disk is expected to form as a result of angular-momentum mixing (the dashed curve in Fig. 5a), while thermally supported gas processed by the feedback and left in the spheroid are too hot and too diffuse to form stars. This corresponds to a classical picture of gaseous components of galaxies in the local Universe. The rate at which the hot gas is accreted into the zone of influence of the BH can be described by the original form of the Bondi accretion (Eq. 59, with $\sigma_{\text{eff}} \approx c_s$):

$$\dot{M}_{\text{B}} = \frac{4\pi G^2 M_{\text{BH}}^2 \rho_{\text{g}}}{c_s^3}. \quad (99)$$

With the approximation that the sound speed $c_s \approx V_c(R_{\text{B}})$, the factor $4\pi G^2 M_{\text{BH}}^2 / c_s^3$ is similar to that of the bursty mode, but the low density ρ_{g} makes the accretion of hot gas much less efficient than the bursty-mode accretion. On the other hand, the rate at which cold gas in the disk is accreted into the zone of influence of the BH can be formulated with an additional factor accounting for the viscosity, as:

$$\dot{M}_{\text{B}} = \frac{4\pi G^2 M_{\text{BH}}^2 \rho_{\text{g}}(R_{\text{B}})}{V_c^3(R_{\text{B}}) C_{\text{visc}}}, \quad (100)$$

where $C_{\text{visc}} = 2\pi [\alpha_{\text{visc}}(H/R)^2]^{-1}$; α_{visc} is the viscosity number; $H/R \ll 1$ is the height-to-radius ratio that describes disk geometry. This formulation was derived by Rosas-Guevara et al. (2015) by comparing the original Bondi timescale and the viscosity timescale, and was adopted as a sub-grid model in the EAGLE simulation (Schaye et al. 2015; Crain et al. 2015). The density ρ_{g} of the cold gas can be large, but the viscosity suggested by them is quite small. This gives a value of $C_{\text{visc}} \sim 10^6$ that largely suppresses the accretion rate of the BH, supporting the assumption made in Paper-I that BH accretion is negligible in a thin galactic disk expected in the slow phase of the halo assembly.

5 MODEL PREDICTIONS AND COMPARISON WITH OBSERVATIONAL DATA

In §§3 and 4, we have presented all the physical prescriptions related to the seeding and growing procedures of SMBHs. Here we join the two procedures together by implementing them numerically into the subhalo merger trees extracted from N-body simulations. As a result, each galaxy in the simulated volume is self-consistently assigned an initial condition (BH seed and the first set of stars formed in the galaxy) according to the matter and radiation environment defined by the host halo and feedback from other galaxies. The initial condition is then evolved along each tree with channels suitable for different evolution stages of the galaxy ecosystem. Predictions for the BH and galaxy populations are thus available throughout the cosmic history in a cosmological volume, and the observational consequences specific to each distinct growth channel can be explored. In the following, we first briefly list the steps of the implementation, and then present the model predictions together with observational implications. A complete description of the implementation, as well as the list of

default parameters adopted, is provided in Appendix A. A number of variants of the model are defined and studied in Appendix C to show uncertainties of the physical prescriptions and to pinpoint the model space to be explored using observations and numerical simulations.

5.1 Numerical implementation of the model

The numerical algorithm takes all dark matter subhalos resolved by an N-body simulation as the input, and extends the merger trees of these subhalos by an EPS-based Monte Carlo algorithm down to a mass of $M_{\text{v}} \sim 10^5 M_{\odot}$ so that the formation of Pop-III stars in mini-halos can be followed (see Fig. 1 for the Pop-III regime in the $M_{\text{v}}-z$ plane). We start from a sufficiently high z , iterate over all snapshots, and populate subhalos in each snapshot with BHs and galaxies by the following steps:

- (i) For each subhalo in the snapshot, we check if a BH seed has already been bred in it. If not, we turn to step (iv) for the seeding procedure. If yes, we take the BH seed and the associated stars as the initial condition, and evolve the BH and the host galaxy by a time interval spanned by the snapshot. Briefly, we determine whether the host halo is in the fast- or the slow-accretion phase, and branch the model accordingly. The key assumption here is the two-phase formation of galaxies proposed in Paper-I (see the summary in §4.1 and Fig. 5): a galaxy in a fast-accreting halo mainly grows the dynamically hot component (bulge), and the BH grows efficiently via capturing low- j sub-clouds formed in the SGC; a galaxy in a slow-accreting halo mainly grows the dynamically cold component (disk), and the BH growth is negligible. For either phase, the algorithm follows the cooling of the gas in the halo, the continuous mode of BH growth and star formation, the metal production, the stellar and AGN feedback, the amounts of gas in different states, and the structure of the gas and stellar components, to produce properties of the galaxy ecosystem. If the subhalo falls into a bigger halo as a satellite, we model it as a close-box, with a declined star formation rate to account for environmental effects. If another subhalo merges into the subhalo in question, a fraction of each mass component (gas, stellar and BH) of the subhalo in the side branch is added to the corresponding component of that in the main branch of the descendant; meanwhile, a fraction of the disk is transferred to the bulge to account for the effects of dynamical heating during the merger.
- (ii) If the host halo makes an temporary excursion to an excessively fast accretion relative to the accretion rate smoothed over a longer timescale (see Eq. 42 for the criterion), we trigger the formation of an SNF nucleus (see §4.2) and solve the nuclear burst using the refinement engine (see §4.5). The Bondi properties and structure of the accretion disk are computed on the fly during the nuclear burst. This bursty mode provides a new channel of star formation and leads to the formation/growth of a compact nuclear star cluster. A wet channel of BH growth is also opened, which supplies gas for a short period of super-Eddington accretion and causes a significant jump in the BH mass if the initial BH mass is small (see Fig. 7 for example events of nuclear bursts). Part of the stars formed in the nucleus subsequently sink to the BH via dynamical interactions, providing a dry channel of BH growth.
- (iii) We model the feedback from the galaxy to the large-scale environment in two forms: a matter form that ejects a fraction of the yielded metal into the IGM; and a radiation feedback that produces the large-scale UV radiation field attenuated by the IGM. These two forms of feedback are self-consistently implemented in the model, as detailed in Appendices A5 and A6. Both forms are critical to determining when and how BH seeds are bred in other subhalos affected by the feedback.

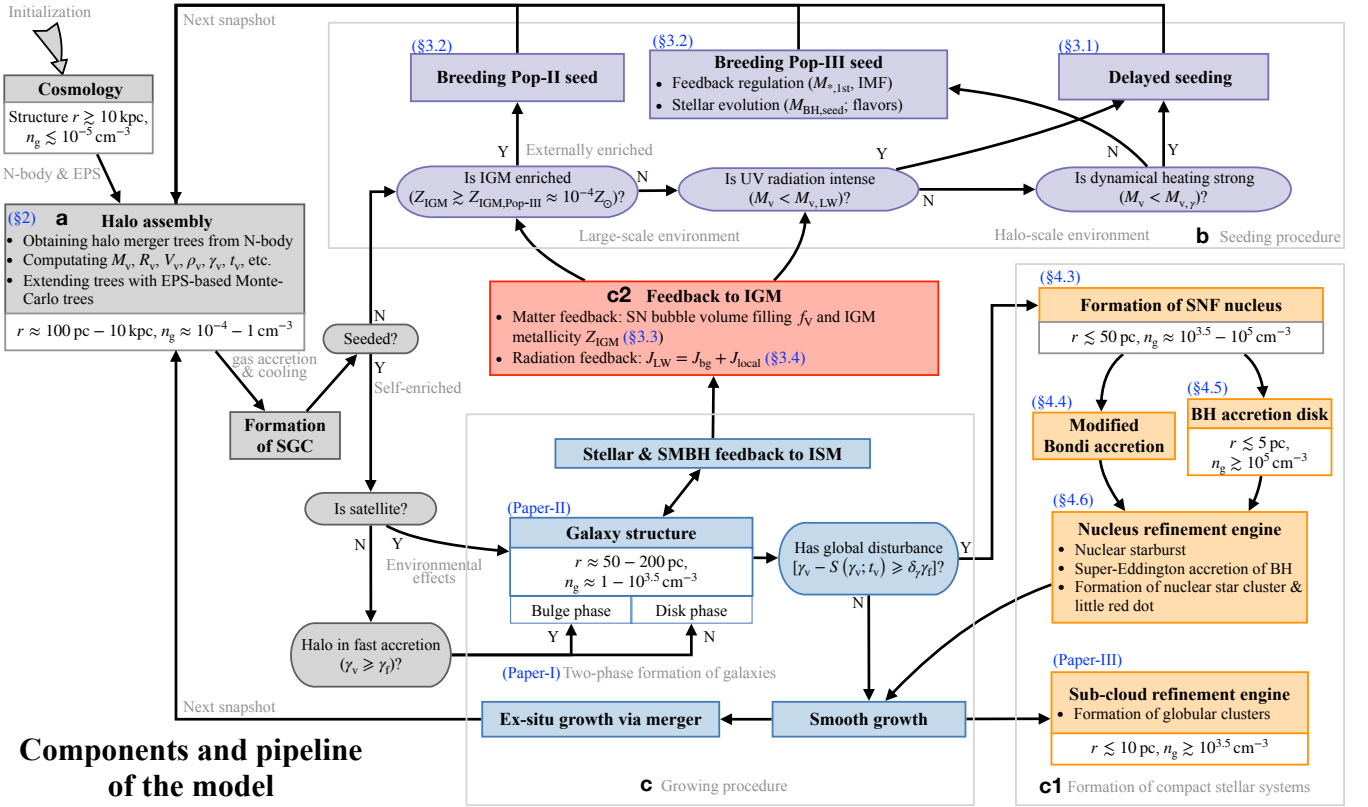


Figure 8. Components and pipeline of the model. The model is built upon the the Λ CDM paradigm, take halos and their assembly histories as input (a), and use two main procedures, the **seeding procedure** (b; see also the diagram in Fig. 2 for a summary of the processes affecting the seeding), and the **growing procedure** (c; see also the diagram in Fig. 5 for a summary of the gas processes), respectively, to seed and grow SMBHs with their host galaxies and halos. Inset c1, sub-procedure for modeling compact stellar systems, include the refinement engine for nuclear bursts and the formation of globular clusters (GCs). Inset c2, sub-procedure for modeling radiation and matter feedback of galaxy formation to the large-scale environment, which provides environmental conditions for the seeding procedure and serves as an interface connecting the two main procedures. Gas densities (n_g) of structures on different scales are listed, taking the progenitors of MW-size galaxies at $z = 9$ (when the Universe is about 10% of its current size) for example (see Fig. 6 for the profile). The structures include the Universe and large-scale structure (LSS), halos, self-gravitating gas cloud (SGC) and galaxy, supernova-free (SNF) nucleus and sub-cloud, and black-hole (BH) accretion disk. The structures will be linked as a ladder-like hierarchy, quantified by sets of equations summarized in Fig. 6, so that one can start from the large scale (a), moving step by step down to the smallest scale. Models of different processes can then be built on top of these structures, and will be extensively discussed in the sections labeled by **blue texts**. See §5.1 for a brief list of the pipeline, and Appendix A for a detailed implementation.

(iv) If no seed has been bred in the subhalo, we check the surrounding IGM of the subhalo to determine whether the metallicity (Z_{IGM}) is high enough to allow the formation of Pop-II stars. If $Z_{\text{IGM}} \geq Z_{\text{IGM,Pop-III}}$, we bred a low-mass BH seed with $M_{\text{BH,seed}} \sim 10 M_{\odot}$ as the descendant of a massive Pop-II star. If the IGM metallicity is too low to produce Pop-II stars, we check if the halo mass has reached the environment-dependent threshold for the formation of Pop-III stars (§3.1.5, Figs. 2 and 3), taking into account the effects of dynamical heating and LW radiation on the collapse of the SGC. If the threshold is reached, a Pop-III star cluster is assumed to form in the SGC, and a BH seed is bred as the remnant of a Pop-III star. To find the mass of the seed, we adopt a radiation-regulated IMF and an optimal sampling method (see Appendix A3) to obtain the masses of the Pop-III stars, and we determine the masses of their remnants by evolving the stellar population specified by the initial mass function (see §3.2 and Fig. 4). The most massive BH remnant is taken as the seed of the central BH. If neither Pop-II nor Pop-III stars can form, the subhalo is kept dark in this snapshot. The ‘flavor’ of the BH seed is defined by the combination of the processes leading to its formation, including the influence of large-scale environment (Z_{IGM} , γ_v and J_{LW}) at the seeding epoch, the level of radiation-

regulation (sub-Eddington versus Eddington) that sets the IMF, and the evolution pathway (CCSN, PISN and CDBH) of the dominant star in the Pop-III cluster that hosts the seed.

Fig. 8 summarizes the numerical algorithm of the model by a flowchart. The flowchart can be roughly divided into three parts: the dark matter (sub)halo properties and merger trees that are used as the input of the model (part a), the seeding procedure (the step iv above) by which BH seeds are bred to set the initial conditions of the galaxy ecosystems (part b), and the growing procedure (the step i above) by which BHs and galaxies are evolved according to the two-phase scenario of galaxy formation (part c). The box c1 represents the formation of the SNF nucleus and the resulted nuclear burst that provides a new channel for BH growth and star formation (the step ii above). The box c2 represents the interface between parts b and c through matter and radiation feedback from the galaxy to the large-scale environment (the step iii above). The modeling pipeline follows all galaxies in the simulated volume together at any given time, so that feedback effects can be taken into account self-consistently when seeding BHs in halos that have not yet formed their first generation of stars. This is different from the previous implementation in Paper-I

(see also Lacey et al. 2016; Spinoso et al. 2023) that takes each merger tree as a processing unit and solves different trees independently.

Once the model is implemented, we can use it to predict the properties of individual galaxies in a cosmological volume. In the following, we show results of the model obtained by applying it to the TNG100-1-Dark simulation (see §2). We focus on statistical properties of the BHs (including seeds) and galaxies predicted. We also use individual examples to illustrate how key physical processes impact model predictions.

5.2 The expected seed populations

5.2.1 Seed mass functions

We define the seed mass function as the number of BH seeds per unit logarithmic mass interval per unit comoving volume formed before a given redshift z :

$$\Phi(M_{\text{BH,seed}}, z) = \frac{1}{V_{\text{u}}} \frac{dN_{\text{seed}}(> z)}{d \log M_{\text{BH,seed}}}, \quad (101)$$

where V_{u} is the comoving volume of the simulation, and $dN_{\text{seed}}(> z)$ is the number of BH seeds that have been bred by z within the mass interval $dM_{\text{BH,seed}}$ in the volume. The black curve enclosing the gray area in Fig. 9a shows $\Phi(M_{\text{BH,seed}}, z = 20)$. The distribution is quite broad, which is produced by the diversity in halo assembly and environmental effects, as discussed below. The peak of the distribution is at $M_{\text{BH,seed}} \approx 500\text{--}1000 M_{\odot}$, close to the mass typically assumed in theoretical and observational studies of Pop-III stars (Schaerer 2002; Wang et al. 2024b). The peak number density is above $0.1 \text{ Mpc}^{-3} \text{ dex}^{-1}$ at $z = 20$, higher than the number density of dwarf galaxies with $M_{*} = 10^8 M_{\odot}$ at $z = 0$ (see e.g. fig. 15 of Chen et al. 2019). This suggests that most galaxies above $10^8 M_{\odot}$ in the local Universe may have their first generation of stars formed and BH seeds bred at $z \gtrsim 20$ (see also §5.2.2 below for the cosmic seeding history). These stars can either remain in the inner galaxy, be dynamically heated to higher orbits, be ejected as halo stars, or merge into other galaxies. Searching for fossils of these stars has been carried out in and around the Milky Way (MW) and has yielded interesting cases with Pop-III-like metallicity (see e.g. Bonifacio et al. 2025, for a review) or with a peculiar PISN-compatible chemical pattern (e.g. Aguado et al. 2023; Xing et al. 2023). Our model suggests that these stars may be used to reveal the formation of the MW back to $z \approx 20$, the epoch when the Universe is younger than 200 Myr.

As discussed in §3.2, the delay in the collapse of SGCs by the heating processes elevates the threshold of M_{v} for the SGCs to form Pop-III stars and increases the amount of gas available to accrete onto the stars. This necessitates a higher luminosity of the dominant star to overcome the accretion flow and regulate the growth of the star itself, modifying the mass spectrum of Pop-III stars at the high- M_{*} end (see Fig. 4a). The effect of the delayed collapse is thus expected to be encoded in the masses of BH seeds. To see this, we divide the Pop-III seeds into two flavors according to whether the dominant star in the progenitor star cluster experienced sub-Eddington or Eddington accretion. The mass functions at $z = 20$ for seeds with the two flavors are shown by the thick blue and thick red curves in Fig. 9a, respectively. Seeds with Eddington progenitors have masses biased towards higher values, compared with those with sub-Eddington progenitor. In particular, a non-negligible fraction of seeds, with an expected number density of $\Phi \gtrsim 10^{-5} \text{ Mpc}^{-3} \text{ dex}^{-1}$, falls in the mass range of the SMS ($M_{\text{BH,seed}} \gtrsim 10^4 M_{\odot}$). SMSs with such high mass are believed to be promising candidates for the progenitors of bright quasars observed at $z \sim 6$ (e.g. Wise et al. 2019;

Latif et al. 2022a). Our model implies that SMSs are sufficiently abundant for such a link, provided that some of the seeds are fed continuously during the fast assembly of their host halos. We will come back to this topic in §5.3.

A prominent feature of the mass function is the bimodality, with an obvious valley at $M_{\text{BH,seed}} \approx 100 M_{\odot}$. This is produced by the evolution of stars in Pop-III clusters. As modeled in §3.2, the dominant star is completely disrupted if it falls in the mass range of a PISN, and the most massive star below the PISN mass limit is expected to result in a BH remnant of $M_{\text{BH,seed}} \lesssim 100 M_{\odot}$ after the CCSN. However, if the dominant star is above the expected mass range of PISN, a DCBH is expected to be produced. Fig. 9b and c show the mass functions for seeds decomposed into three sub-categories according to the evolution (CCSN, PISN, DCBH) of the dominant star, for sub-Eddington and Eddington cases, respectively. In both cases, the masses of the seeds with CCSN and PISN flavors are comparable, with the distribution peaked below $100 M_{\odot}$, and seeds with the DCBH flavor have much higher mass, with the distribution peaked at about $500 M_{\odot}$. As we will show in §5.3, a galaxy is expected to experience a short period of quiescent phase immediately after a seed is bred, because the associated AGN feedback strongly suppresses the star formation and BH growth. Since BHs with higher mass are more efficient in feedback (in terms of the energy released per unit accreted mass), seeds with the DCBH flavor may quench the galaxy for a longer period of time, while seeds with the CCSN and PISN flavors may allow the galaxy to rejuvenate earlier. The bimodal distribution of the seed mass is thus expected to produce a bimodal distribution in the age of stars after the seeding.

Both the UV radiation and fast accretion of halo are modeled in this paper to account for environmental processes that ‘heat’ the SGC and delay its collapse (§§3.1.3 and 3.1.4). Although both factors are expected to produce Pop-III seeds with higher mass, their relative importance differs in different SGCs. Dynamical heating is stronger if the halo assembly is faster, and is thus expected to be more important for halos of higher mass or at higher redshift (see Fig. 3b). The effect of the LW radiation is stronger if galaxies with higher SFRs are present in the neighborhood, and is thus expected to be more important at lower redshift when a larger number of massive, star-forming galaxies have formed in over-density regions. To see the difference of these two factors, we select BH seeds bred in (i) ‘fast halos’, defined as those with $\gamma_{\text{v}} \geq 3$ at the seeding epoch; (ii) ‘LW halos’, defined as those with $J_{\text{LW},21} \geq 7.5$. The thresholds are chosen so that the expected halo mass for the collapse of SGC is about $10^7 M_{\odot}$ (see Fig. 3). The two orange curves in Fig. 9a show the mass functions of BH seeds bred in the two types of halos, respectively. Seeds in both types of halos are dominated by Eddington accretion, highlighting the importance of the large-scale environment in modulating when and how BH seeds are bred. The masses of seeds in LW halos are higher than those in fast halos, consistent with the expectation that the LW radiation is more efficient at lower z when subhalos that have not been seeded with BHs reach higher masses.

For comparison, we show $\Phi(M_{\text{BH,seed}}, z = 0)$, the seed mass function at $z = 0$, in Fig. 9d–f, for all seeds and those with different flavors as defined above. The mass function at $z = 0$ is similar to that at $z = 20$ in terms of the normalization, the bimodal shape, and the contributions of different flavors. A noticeable difference is that the mass function at $z = 0$ is broadened towards both the low and high mass ends, and become more prominent at $z = 0$. Both tails are dominated by LW halos and fast halos, implying that both of the two environmental factors are important in delaying the seeding. LW radiation and dynamical heating now have a *divergent effect* on the masses of BH seeds. On the one hand, it delays the collapse of an

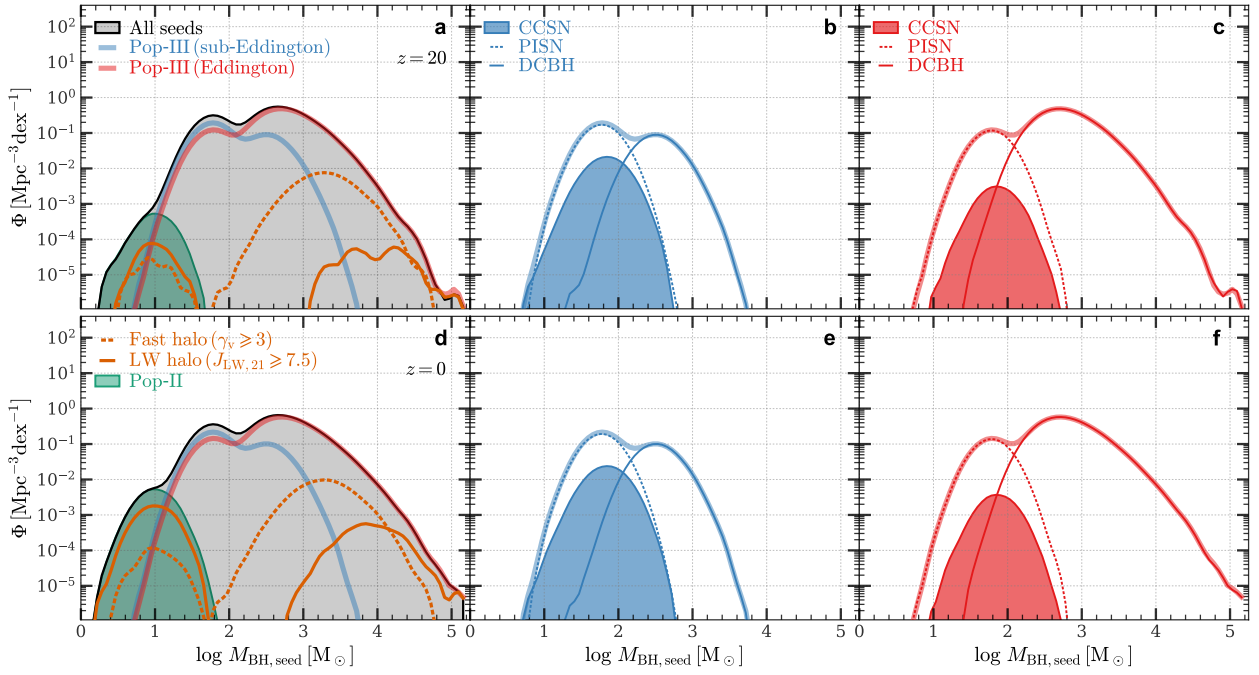


Figure 9. Mass functions of BH seeds. **a**, $\Phi(M_{\text{BH,seed}}, z = 20)$, mass functions of all seeds that have been bred until $z = 20$ in the simulation volume (shaded area bound by **black** curve) and of seeds with different ‘flavors’ (selected by different criteria): descendants of Pop-III stars experiencing sub-Eddington (**blue**) and Eddington (**red**) accretion, those born from dynamically heated SGC in fast-accreting halos with $\gamma_v \geq 3$ (**orange dashed**) and from LW-irradiated SGC (**orange solid**), and descendants of Pop-II stars fed by enriched IGM (**green**). Each category of the Pop-III descendants is further divided into three sub-categories according to the stellar evolution that leads to the BH seed (CCSN, PISN and DCBH) for the dominant star in the host Pop-III star cluster. **b** and **c** show the mass functions of these sub-categories. **d–f** are similar but include all seeds bred until $z = 0$. The mass spectrum of BH seeds is quite broad and exhibits two main modes, suggesting that galaxy bimodality has already appeared at $z \approx 20$. Environmental factors have a divergent impact on the masses of BH seeds, producing both high-mass and low-mass bumps in the mass function, with DCBH and Pop-II flavors, respectively. See §5.2.1 for details.

SGC, so that more massive DCBHs are expected to be bred owing to the higher gas mass available for the formation of Pop-III stars (red thin curve in panel f). On the other hand, it gives more time for SNe from nearby galaxies to enrich the mini-halo, so that Pop-II seeds are more likely to form via efficient cooling provided by the metals (green shade in panel d). In fact, most Pop-II seeds are bred at $z \lesssim 20$, in contrast to Pop-III seeds that are mostly bred at $z \gtrsim 20$ –30 (see also §5.2.2 and Fig. 10a). The interplay between dynamical heating and the two forms of feedback (matter and radiation) from galaxy formation to the large-scale environment may thus leave interesting imprints on the properties of the seed population.

5.2.2 The cosmic seeding history

To quantify the redshift distribution of seed formation, we define the cosmic seeding rate density as the number of BH seeds bred per unit logarithmic cosmic scale factor per unit comoving volume:

$$\frac{dn_{\text{seed}}}{d \log(1+z)} = \frac{1}{V_u} \frac{dN_{\text{seed}}(>z)}{d \log(1+z)}. \quad (102)$$

The cumulative value of the cosmic seeding rate density,

$$n_{\text{seed}}(>z) = N_{\text{seed}}(>z)/V_u, \quad (103)$$

is referred to the cosmic cumulative seeding density. This mimics the definition of the cosmic star formation history (Madau & Dickinson 2014), but is specific to the BH seeding.

In Fig. 10, the top row shows $dn_{\text{seed}}/d \log(1+z)$ as a function of redshift for seeds of different flavors as defined in §5.2.1, while the bottom row shows the corresponding cumulative distributions. The

total seeding rate density rises steeply with time at $z \gtrsim 25$, reaches the peak at around this redshift, and then declines rapidly to less than 10^{-2} (10^{-3}) $\text{Mpc}^{-3} \text{dex}^{-1}$ at $z = 10$ ($z = 5$). The low rate at $z \lesssim 10$, as well as the low mass of the Pop-III star clusters, implies that the searching for ongoing formation of Pop-III stars and BH seeds are challenging even with powerful telescopes like JWST. Cosmic dimming at $z \approx 25$ also challenges the detection, even though these objects were actively forming at that epoch. The first generation of seeds are remnants of Pop-III stars with sub-Eddington accretion, followed by the domination of the remnants of Pop-III stars with Eddington accretion at $z \lesssim 30$. Seeds with sub-Eddington accretion are more likely to have PISN flavor, while those with Eddington accretion are mostly DCBHs. Most seeds have already been bred by $z \sim 20$, consistent with the expectation from seed mass functions (§5.2.1 and Fig. 9).

The two environmental factors, fast halo assembly and LW radiation, play an important role in delaying the seeding at high and low redshifts, respectively. The relative importance crosses over at $z \approx 15$ –20, as seen from the evolution in the seeding rate density of the fast halos and LW halos. Most seeds bred at $z \lesssim 10$ are delayed by the LW radiation, highlighting the critical role of LW radiation in the production of late-formed Pop-III stars and BH seeds, which are targets of some recent JWST projects (e.g. Matthee et al. 2023; Naidu et al. 2024; Fujimoto et al. 2025; Hsiao et al. 2025) and future surveys. Some of the seeds bred at $z \lesssim 20$ are low-mass Pop-II remnants, reflecting the divergent effect of heating sources on the mass of BH seeds, as discussed in §5.2.1.

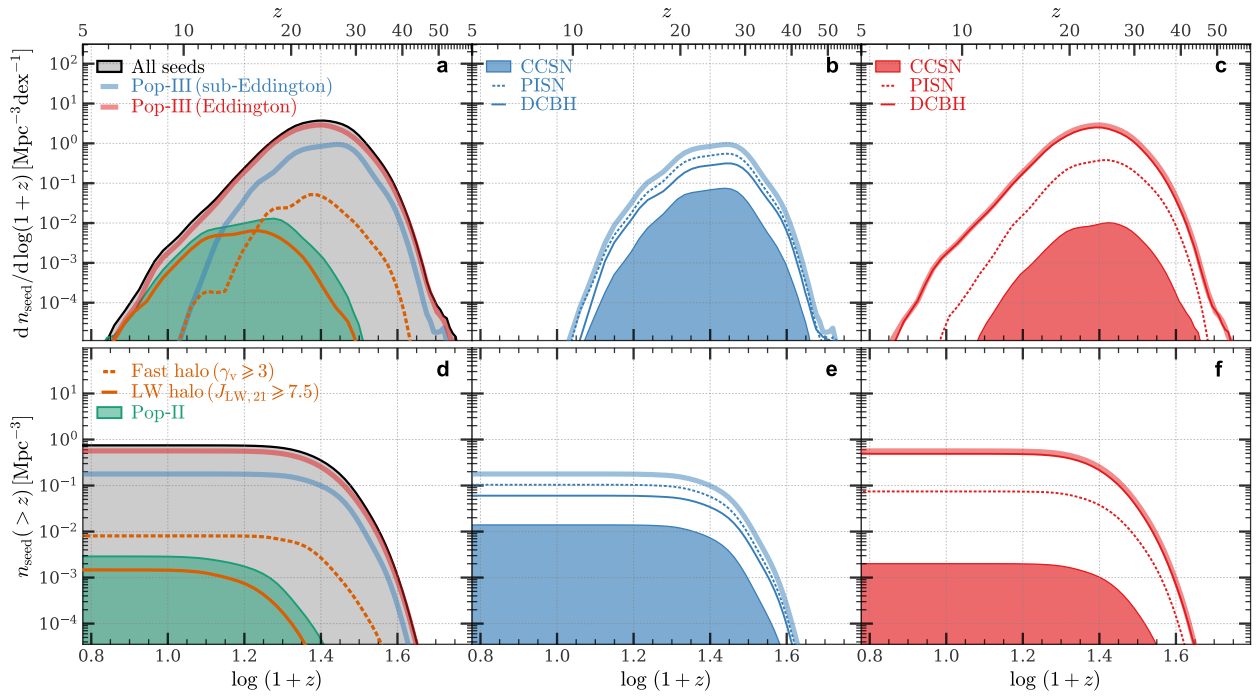


Figure 10. Cosmic BH seeding history. Redshift distribution of BH seeds with different flavors. **a–c**, cosmic seeding rate density, (the number of BH seeds bred per unit logarithmic scale factor per unit comoving volume) as a function of z . **d–f**, cosmic cumulative seeding density (the number of BH seeds bred before z) as a function of z . The results are shown for all seeds (**black** curve) and for seeds with different ‘flavors’. Pop-III seeds with sub-Eddington (**blue**) and Eddington (**red**) accretion are further divided into sub-categories according to the stellar evolution that leads to the seed (CCSN, PISN and DCBH). At $z \approx 20$, most seeds have already been bred. Pop-II seeds form at systematically lower redshift than Pop-III seeds, and dominate the seed population at $z \lesssim 10$. See §5.2.2 for details.

5.2.3 The seeding atlas: places and environments of sowing

The significant effects of the environment on the breeding of seeds highlight the importance of the explicit modeling of the environment to identify when and how seeds in individual mini-halos are bred. One critical part is that the conditions for a seed to be bred may not be deduced solely from the internal properties of the host (sub)halo. This complexity can be understood by considering two extreme cases. An extremely massive halo (e.g. $M_{v,z=0} = 10^{15} M_{\odot}$) often resides in an over-dense region where the LW radiation and the IGM enrichment would be strong if many halos in this region have already started forming stars. However, a halo with such a high mass is rare, and is likely to be the first one in the region to reach the H_2 -cooling threshold. LW radiation and IGM enrichment are thus likely to be completely absent for the halo, leaving the dynamical heating as the only mechanism to delay the seeding. As the other extreme, a low-mass halo (e.g. $M_{v,z=0} = 10^{10} M_{\odot}$), if formed in an over-dense region, can be strongly irradiated by the LW radiation from nearby star-forming galaxies and be bred with a high-mass Pop-III seed. However, if the gas environment of the halo has already been polluted with metals, a low-mass, Pop-II seed may be bred instead. Since metal diffusion is much slower than the propagation of radiation, small halos residing on the two sides of the boundary separating pristine and enriched regions around a massive and LW-emitting galaxy can have a diverse seed population. Thus, the location of a halo and the large-scale environment around the location are necessary input to the seeding procedure. This also supports the use of the global integrator, instead of the tree-based processing unit, in the modeling pipeline (§5.1).

To see how the seeding processes in individual mini-halo are affected by the environment, Fig. 11 shows the ‘seeding atlas’, i.e.

the distribution of BH seeds in the M_v - z plane. Seeds in individual mini-halos are represented by dots, color-coded by the seed mass ($M_{BH,seed}$; panel a) or one of the three environment properties evaluated at the seeding epoch: the IGM metallicity (Z_{IGM} ; panel b), the LW radiation intensity ($J_{LW,21}$; panel c), and the specific halo growth rate (γ_v ; panel d). For reference, we also show the median assembly histories of halos with different masses at $z = 0$ and the regime in the M_v - z plane where Pop-III stars are expected to form (see Fig. 1 for details). With these maps, the diverse populations of seeds that can be bred in a halo owing to the difference in environment, as well as the M_v and z at which the seeds are bred, can be visually identified by the intersection of the halo assembly history and the distribution of the seeds in the M_v - z plane. An empirical strategy of seeding to be applied to, e.g. hydro simulations, may also be constructed from the atlas, to partly account for the statistical expectation and variation of the seed population at a given halo mass, if the environmental information is unavailable.

The emergence of the first generation of stars and BH seeds is at $z \gtrsim 40$ when the progenitors of the extremely massive halos with $M_{v,z=0} \gtrsim 10^{14} M_{\odot}$ reach the H_2 -cooling threshold (black and dark blue dots in the lower-right corner of Fig. 11a). IGM has not yet been polluted by star formation in this era, and the first generation of stars are thus all Pop-III with nearly primordial chemical composition (panel b). LW radiation field has yet to be built up at this epoch (panel c), and dynamical heating is the main mechanism that delays seeding (panel d). As shown in §5.3.1 and Fig. 14, the assembly of a halo fluctuates significantly in the fast phase. A mini-halo with a temporary excursion to $\gamma_v \ll \gamma_{crit}$ (Eq. 22) thus gets a chance to escape the dynamical heating and form Pop-III stars. Our model thus predicts that the places for the formation of first generation of stars and BH seeds in the Universe are the progenitors of the

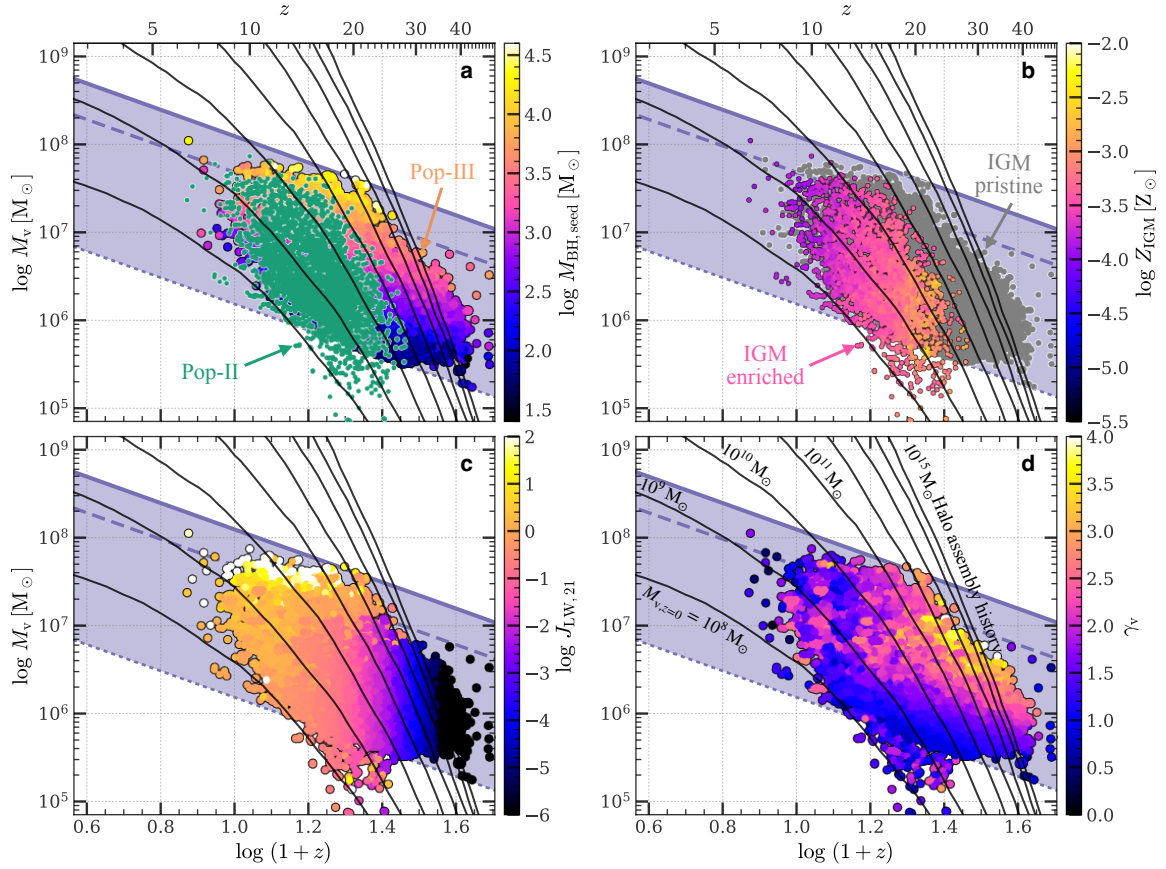


Figure 11. The atlas of BH seeds. Here we show the distribution of BH seeds in the M_v - z plane (host halo mass versus redshift at the seeding epoch), color-coded by the mass of the BH seed ($M_{\text{BH,seed}}$, **a**), metallicity of IGM surrounding the subhalo (Z_{IGM} **b**), intensity LW radiation irradiating the subhalo ($J_{\text{LW},21}$, **c**), or specific halo growth rate (γ_v , **d**), all evaluated at the seeding epoch. In **a**, green dots represent Pop-II seeds, while the others represent Pop-III seeds. In **b**, grey dots represent those embedded in pristine IGM ($Z_{\text{IGM}} = 0$), while the others are color-coded by Z_{IGM} . All seeds are color-coded in **c** and **d**. All seeds bred in the simulation volume are included in this plot. Kernel smooth is applied to the colored properties to clarify the trends. Each black curve shows the median assembly history of halos with a given mass at $z = 0$ (from 10^8 to $10^{15} M_\odot$, each separated by 1 dex), and purple curves and band indicate the regime where Pop-III stars are expected to form and seed the BHs (see Fig. 1). A few Pop-III seeds are outside the Pop-III regime because of the numerical discretization of the halo assembly histories. This figure illustrates when, where and how BH seeds are bred in different halos and how environmental processes affect the formation of the first generation of stars and BH seeds. See §5.2.3 for details.

extremely massive halos that reach the H_2 -cooling limit the first, are most likely to appear in over-dense regions, are embedded in a pristine IGM, and have temporary excursions to low γ_v . The sizes of individual SGCs at such high redshifts are only a few pc (see the texts below Eq. 14); together with the high density threshold required for the low-metallicity gas to fragment, we expect that the Pop-III star clusters are more compact than the Pop-I/II globular clusters formed later (see the red and blue regions in Fig. 1a). Due to the low masses of the SGCs expected from the low masses of the H_2 -cooling halos, the masses of the dominant stars in the first set of Pop-III star clusters are less than a few hundred M_\odot . CCSNe and PISNe are expected to end the lives of these stars, leaving BH remnants with masses $\lesssim 100 M_\odot$ in the clusters. In some cases where mini-halos form very early and remain at high γ_v , DCBHs with masses greater than a few hundred M_\odot can be bred at $z \gtrsim 40$ (purple dots in the lower right corner of panel a).

If a massive halo can sustain a high γ_v over a long period after the halo reaches the H_2 cooling threshold, seeding can be significantly delayed until the halo grows above a higher value of the collapsing threshold, $M_{v,\gamma}$ (Eq. 23). The consequence is that a larger amount of gas can be converted into stars and a more massive seed can be

bred. The cloud of points at $M_v \approx 10^7 M_\odot$, $z \approx 30$, with purple-to-yellow colors in Fig. 11d represent this population of seeds bred in ‘dynamically heated’ halos. For a massive halos with $M_v \gtrsim 10^{13} M_\odot$ at $z = 0$, the seeds are mostly bred at $z \gtrsim 20$, with mass spanning a wide range from a few tens to about $10^4 M_\odot$ (panel a), with flavors being either sub-Eddington or Eddington in terms of the accretion rate during the formation of the dominant star, and being CCSN, PISN and DCBH in terms of the evolution of the dominant star. The contribution of heating from LW radiation is expected to be unimportant for these halos because their seeding epochs are too early for the radiation field to build up (panel c). Note that the assembly histories and environments of massive halos are quite diverse (see e.g. §4.2 of Chen et al. 2023b), and therefore the seeds of individual halos can deviate from those expected for the median assembly history.

Less massive halos with $M_{v,z=0} \lesssim 10^{13} M_\odot$ follow a trend similar to that of the massive ones, albeit with some shift in the M_v - z plane: a halo above the H_2 -cooling threshold is bred with a Pop-III seed if the assembly history slows down to $\gamma_v \lesssim \gamma_{\text{crit}}$; a sustained high γ_v delays the seeding and produces a more massive seed. Since a less massive halo reaches a given mass threshold at lower redshift and since the mass threshold at a given virial temperature

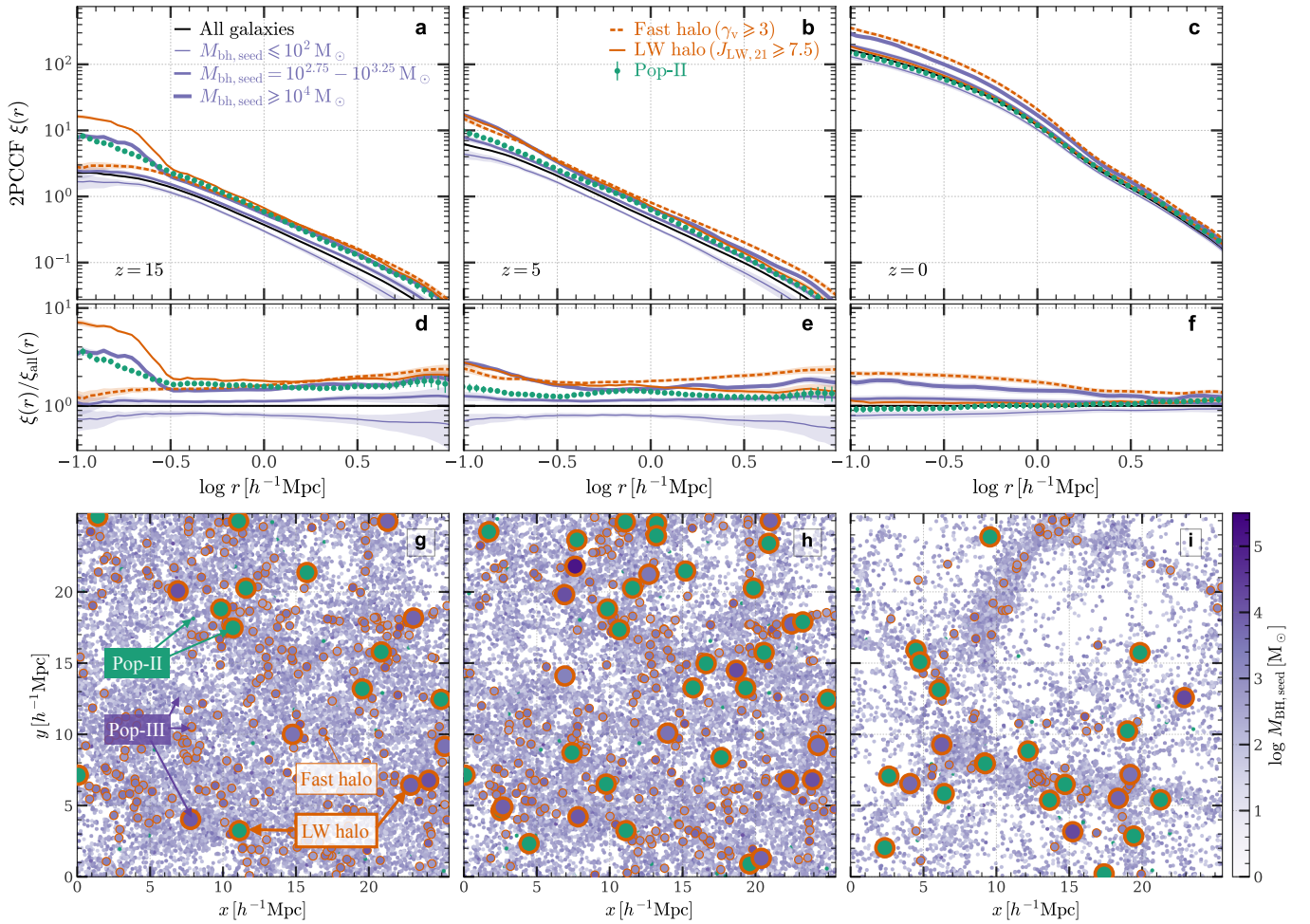


Figure 12. Spatial distribution of galaxies that bred different BH seeds. Three columns show the results at $z = 15, 5$ and 0 , respectively. **a–c**, 2PCCFs (ξ) as functions of comoving separation r for different target samples of galaxies. At each given z , the 2PCCFs of the target samples are measured with respect to a common reference sample that includes all galaxies at z . The target samples displayed are: all galaxies (black; equivalent to the 2PACF of the reference sample; denoted as ξ_{all}); galaxies with Pop-III seeds that have different seed masses (purple; $M_{\text{BH,seed}}$, the mass of BH at the breeding epoch); galaxies hosted by fast halos (orange dashed); galaxies hosted by LW halos (orange solid); galaxies with Pop-II seeds (green). **d–f**, relative bias of each target sample, defined as the ratio between the 2PCCF of the sample and ξ_{all} . Curve with error bars (or shade) is the mean with standard deviation evaluated from bootstrap resampling. **g–i**, dots represent galaxies in the same slice with a thickness of 5 Mpc at the three redshifts, respectively. Style of a dot is encoded according to the seed bred in the main branch: purple if the seed is a Pop-III remnant (more massive seed is darker); green if the seed is a Pop-II remnant; with orange edge if the seeding is in a fast (small size) or a LW halo (big size). This figure shows that BH seeds with different flavors exhibit distinct clustering patterns, and some of the patterns are persistent over cosmic time. See §5.2.3 for details.

is higher at lower redshift (see Fig. 1a), the distribution of the seeds bred with a given $M_{\text{BH,seed}}$ or bred in halos with a given γ_v shifts towards the upper-left corner in the M_v - z diagram as $M_{v,z=0}$ decreases. Dynamical heating thus defines a regime in the atlas that has a fusiform area, filled with tilted, stripe-like equal- $M_{\text{BH,seed}}$ or equal- γ_v contours (panels a and d).

An additional feature in the seeding environments of low-mass halos is the presence of strong LW radiation. Such a dependence of LW radiation on halo mass is a consequence of the combination of three factors: (i) the cosmic SFR density increases with time before $z \approx 2$ (Madau & Dickinson 2014; see also fig. 16 of Paper-I for the predictions of our model), implying a stronger cosmic LW background at lower redshift; (ii) the modulation of local LW radiation by the IGM is less efficient at lower redshift, as detailed in Appendix A5, enabling more effective propagation of the LW radiation from sources to destination halos; (iii) halos with lower masses reach the Pop-III regime (purple bands in Fig. 11) later than those with higher masses.

Low-mass halos under effective LW radiation can thus have significantly delayed seeding times and significantly elevated seed masses. The orange-to-white dots in Fig. 11c represent this LW-illuminated regime of seeding, which forms a tip-like structure in the M_v - z diagram. Seeds produced in this regime have the latest seeding times and the highest halo masses among all Pop-III seeds. Halos with $M_v = 10^{12} M_\odot$ marginally intersect with the LW-illuminated tip and halos with lower mass intersect more, as expected from the above three factors that cause the halo mass dependence.

Another feature in the seeding environments of low-mass halos is the presence of enriched IGM. Similar to the LW radiation, the enrichment level of the IGM increases with time, since metal yield and diffusion are both powered by star formation. The difference between metal diffusion and radiation propagation is that the former is much slower. As a result, the enriched gas environments only appear in the seeding in halos with lower mass than those affected by the LW radiation. Fig. 11b shows that most halos with enriched

gas environments at the seeding epochs have $M_V \lesssim 10^{11} M_\odot$. The cloud of green dots in panel a also shows that only a fraction of them have Z_{IGM} higher than the threshold of $Z_{\text{IGM,Pop-III}} = 10^{-4} Z_\odot$ set for breeding Pop-II seeds. The distribution of Pop-II seeds in the atlas appears as an elongated stripe, with two separate components pulled off by the two sources of heating, respectively. The first component is at $z \approx 10\text{--}20$, with M_V right above the H_2 cooling threshold, formed in the area that partly overlaps the fusiform area of dynamic heating. The other component overlaps the LW-illuminated tip, with a halo mass close to the H-cooling threshold, bred at $z \lesssim 15$, with some cases extending to $z \approx 5$. For both components, low-mass halos with $M_V \lesssim 10^{10} M_\odot$ dominate the Pop-II seeds, as expected from their late formation times.

In summary, the seeding atlas can be divided into three distinct regimes according to the dominating environmental processes that affect the seeding:

- (i) **The regime of dynamical heating**, which spans a fusiform area in the M_V - z plane. Stronger and more persistent dynamical heating leads to a longer delay of the seeding, and produces a more massive seed.
- (ii) **The regime of LW radiation**, a tip-like structure which extends the low- z , high- M_V end of the regime of dynamical heating. Stronger and more persistent LW radiation causes a longer delay of seeding and produces a more massive seed.
- (iii) **The regime of enriched IGM**, which is a two-component stripe on top of the low- z parts covered by the above two regimes.

The three environmental processes take effects via mediators that are different in the domain of influence. Thus, mini-halos affected by each of the processes are expected to have a distinct clustering pattern. Such a pattern encodes the domain of influence of the process when the seeds are bred in the mini-halos. Owing to the preservation of the matter distribution on large scales, the clustering pattern is expected to be persistent, and may be retrieved later to obtain the archive of the seeding histories for the halos that have grown massive enough to host observable galaxies.

To quantify the clustering pattern of a given target sample of galaxies at a given redshift z , we measure the two-point cross-correlation function (2PCCF) of the sample with respect to a reference sample. The 2PCCF is defined as the expected excess of number density, relative to the mean number density in a representative volume, of the galaxies in the reference sample at a given distance r from a galaxy in the target sample. Numerically, the 2PCCF can be estimated as

$$\xi(r) = \left\langle \frac{n_{\text{ref}}(r)}{\bar{n}_{\text{ref}}} \right\rangle_{\text{target}} - 1, \quad (104)$$

where $n_{\text{ref}}(r)$ is the number density of galaxies in the reference sample falling in a spherical shell of a finite thickness around the distance r from a galaxy in the target sample, \bar{n}_{ref} is the mean number density of galaxies in the reference sample, and the average is taken over all galaxies in the target sample. Note that we adopt the 2PCCF, instead of the two-point auto-correlation function (2PACF), because the former is more stable for small target samples such as that hosting Pop-II seeds, provided that the reference sample is sufficiently large. With our choices, the 2PCCF of the reference sample, which we denote as ξ_{all} , is the 2PACF of the reference sample itself. The ratio between any ξ and ξ_{all} , either evaluated at a given r or averaged over a range of r , is referred to as the relative bias at the corresponding spatial scale (e.g. Zhang et al. 2025a).

Here we measure the 2PCCFs of various target samples of galaxies at a given z selected according to the masses and flavors of the seeds that have been bred in the main branches of the subhalos hosting the galaxies. A common reference sample consisting of all galaxies at

z is adopted to maximize the statistical stability. In Fig. 12a–c, we show $\xi(r)$ of the target samples at $z = 15, 5$ and 0, respectively. The corresponding results for the relative bias are shown in panels d–f. For a demonstration, panels g–i show the locations of individual galaxies using scatters with styles marking the mass and flavor of the seed.

For galaxies with Pop-III seeds, the 2PCCF is higher at all scales if the seeds are more massive (purple curves in Fig. 12a–c). At $z = 15$, the 2PCCF of galaxies with $M_{\text{BH,seed}} \geq 10^4 M_\odot$ (mostly DCBHs with Eddington-accreting progenitor stars; see Fig. 9a) is a factor of 2–4 higher than that of galaxies with $M_{\text{BH,seed}} \lesssim 10^3 M_\odot$ (mostly remnants after CCSNe and PISNe). The difference in the 2PCCF between the target samples with different $M_{\text{BH,seed}}$ becomes less and less pronounced over time. As shown in Fig. 10a and d, most seeds have already been bred by $z \approx 15\text{--}20$, and the cosmic seeding rate density is negligible at $z \lesssim 5$. Thus, the decline of the $M_{\text{BH,seed}}$ -dependence of the 2PCCF at $z \lesssim 15$ is due to the evolution of the cosmic large-scale structure rather than the cosmic seeding activity. Panels g–i clearly show that galaxies are less clustered at $z = 15$, but become more clustered later as structures form.

As shown by the orange and green curves in Fig. 12a–f, halos affected by each of the three environmental processes have a distinct clustering pattern at a given redshift and such a pattern shows a distinct trend of evolution. Fast halos (with $\gamma_V \geq 3$ at the seeding epoch) at $z = 15$ has a relative bias of about 2–3, quite independent of the scale at $r \geq 0.3 h^{-1} \text{Mpc}$. At $z = 5$, the relative bias is comparable to that at $z = 15$ at $r \geq 0.3 h^{-1} \text{Mpc}$. Such a high relative bias reflects the requirement of an excessively high γ_V , and thus an excessively over-dense environment, in comparison to galaxies in normal halos, for the seeding to be significantly delayed. Interestingly, the relative bias of fast halos at small scale ($r \approx 0.1 h^{-1} \text{Mpc}$) shows a clear increase with time and reaches about 2–3 at $z \lesssim 5$. The build-up of such a ‘one-halo’ term (e.g. Jing et al. 1998; Peacock & Smith 2000; Zheng et al. 2005) in the clustering pattern reflects the large number of satellite subhalos assembled into the host halo during the fast accretion. At $z = 0$, the relative bias of fast halos at large scale declines due to the cease of cosmic seeding activity and the evolution of the large-scale structure (seen from the decline of the number of fast halos in panels g–i), similar to the evolution of clustering of galaxies bred with massive DCBHs. The small-scale upturn of the relative bias is persistent and extends to scales up to a few Mpc, reflecting the continuation of halo growth and the assembly of satellite subhalos.

The clustering strength of LW halos (with $J_{\text{LW},21} \geq 7.5$ at the seeding epoch) at $z = 15$ is strong at the scales displayed in Fig. 12a. The relative bias is about 2 at $r \geq 1 h^{-1} \text{Mpc}$, and reaches about 7 at $r \approx 0.1 h^{-1} \text{Mpc}$. At such high redshift, the global LW radiation field has not been established yet, and strong LW radiation is only available from nearby galaxies with intense star formation (see Appendix A5). The strong clustering at large scales is thus expected since the production of LW photons requires the presence of galaxies with intense star formation, which is more likely to happen in over-dense regions harboring early halo formation. The excessively strong clustering at small scales reflects the properties of the domain of influence of the LW radiation. Since the local term of LW radiation declines rapidly with distance as r^{-2} (see Eq. A23), Halos affected by LW radiation (LW halos) tend to have close neighbors and thus enhanced small-scale clustering. Note that we have only seen a small number of LW halos in the slice of panel g, suggesting that the chance to get such a close pair of galaxies is small at $z = 15$ (see also Fig. 10d for n_{seed}). Towards lower redshift, the global LW background gradually builds up, and the requirement of close neighbors for LW halos is relaxed. This can be seen from the decline of the small-scale upturn of the

relative bias of LW halos at $z = 5$ and 0, and the less pronounced large-scale clustering of LW halos than that of fast halos at these redshifts. The large-scale clustering of LW halos declines with time, again due to the mergers of halos, as seen from the decline of the number of LW halos from panels h to i.

Similar to fast halos and LW halos, the large-scale clustering of galaxies with Pop-II seeds (green in Fig. 12) is higher than ξ_{all} at $z \gtrsim 5$, implying that Pop-II seeds are embedded in over-dense regions. This is consistent with the aforementioned finding that the two components of Pop-II seeds shown in Fig. 11a overlap the regimes of dynamical heating and LW radiation, respectively, and that both factors preferentially select over-dense regions for the seeding. At $z = 15$, the small-scale clustering of galaxies with Pop-II seeds is exceptionally strong, as is expected from the typical bubble sizes of star-forming galaxies that enrich the IGM with metals (see Eq. A28 and Fig. A3). Towards lower redshift, the relative bias of galaxies with Pop-II seeds declines at small scales, and even becomes slightly lower than that of the reference sample at $z = 0$. Such a distorted pattern of clustering is caused by the competition between two factors, one is that the generation of Pop-II seeds requires a crowd of galaxies to enrich the IGM with metals, and the other is that Pop-II seeds tend to be bred in low-mass halos (Fig. 11) that have a small number of satellite subhalos. At $z \lesssim 5$, the evolution of the clustering of galaxies with Pop-II seeds again follows that of the large-scale structure.

5.3 The growth of galaxies and black holes

Once a BH seed is bred in a subhalo, we follow the growing procedure (see Fig. 8c for the pipeline) to model the subsequent growth of the galaxy ecosystem. Properties of the BH and the host galaxy are predicted at each snapshot, and the evolution histories of these properties are obtained by linking all the snapshots along the subhalo merger tree.

One important subject in understanding BH growth is to identify the co-evolution pattern of BHs and their host galaxies (e.g. Bower et al. 2017; Habouzit et al. 2021; Li et al. 2024). Models implementing different assumptions for the physical processes driving the co-evolution are expected to produce different patterns, which can be targeted by observations to test model assumptions.

Two key processes incorporated in this paper for the early evolution of BHs and host galaxies are the seeding (the formation of the first generation of stars and the breeding of BH seed) and the growth of both the BH and galaxy in the bursty mode associated with the formation of an SNF nucleus. The former sets the initial condition for the co-evolution, while the latter drives the evolution onto a pathway that eventually converges to the relation between BHs and galaxies observed in the local Universe (Kormendy & Ho 2013; Greene et al. 2020; Graham & Sahu 2023) when the growth of BHs and galaxies becomes more steady. Here we focus on the build-up of the BH mass and stellar mass of each galaxy and to study how the processes implemented in our model shape the relation between the two masses. In §6, we discuss other potential observational signatures to be explored in the future.

5.3.1 Growth channels

As detailed in §4, our model incorporates three modes for the growth of a BH and its host galaxy after the seeding:

- (i) **The bursty mode**, in which an SNF nucleus forms in the galactic center, the BH experiences a super-Eddington accretion (referred to as the wet channel in the bursty mode) if it is not too massive, and

stars form in a compact nuclear star cluster. The requirement for triggering the bursty mode is a global disturbance that raises the dynamical hotness of the gas, which we capture by the temporary excursion of the halo assembly to a high γ_v (Eq. 42). The short dynamical timescale of the SNF nucleus, about 1 Myr, defines the bursty nature of the growth in this mode (see Fig. 7 for example). Dynamical interactions of stars may also contribute to the growth of the BH in this mode, which we refer to as the dry channel.

- (ii) **The continuous mode**, in which the BH captures the ballistically moving sub-clouds of low j born from the SGC, and stars form within each sub-cloud. Due to the large number of sub-clouds, star formation and BH growth in this mode is less bursty. BH growth in this mode is expected to be significant if the galaxy is in a dynamically hot state that can be sustained by the fast accretion of the halo over a long timescale ($\gtrsim t_v$), and to be vanishingly small if the galaxy is in a dynamically cold state after the halo accretion can no long impact the gravitational potential (see the discussion below Eq. 100).
- (iii) **The merger**, in which two galaxies and their central BHs merge to form a new galaxy and a new central BH.

The regimes in which the above modes dominate the BH growth and star formation in galaxies are expected to be different. The bursty mode requires a global disturbance to generate the hotness of the gas. This is more likely to happen if the merger rate of the host halo is high, and is thus expected to be more common in the early Universe and in a halo with higher mass. The accretion rate of a BH in the bursty mode is stronger if M_{BH} is smaller, so that the Eddington ratio is initially higher and the feedback regulation is less effective, or if M_v of the host halo is larger, so that the SNF nucleus is more massive to supply more gas to the BH (see texts below Eq. 69 for an estimate). Thus, the growth of a BH through the bursty mode is expected to be more significant in the early Universe (before the BH has grown massive) or in a halo with higher mass. Similar conclusion applies to the star formation in the bursty mode, since a less massive BH is less competitive compared to star formation (see §4.5.1), and a more massive halo can supply a more massive nucleus to form stars. However, the continuous mode depends on the fast accretion of the host halo and the global properties of the SGC. Such a mode is thus expected to be efficient during the entire fast phase of the halo, even if the SNF nucleus is not formed or is negligibly small. Finally, the contribution from merger dominates if other channels are not active. This is more likely in later times and/or in a halo with higher mass where galaxy quenching suppresses the growth in the other modes.

To see the contribution of each channel to the growth of BHs and their host galaxies, Fig. 13 shows the evolution of individual galaxies in the $M_{\text{BH}}-M_*$ plane, separately for different halo masses at $z = 0$ ($M_{v,z=0}$). In each range of $M_{v,z=0}$, we pick a representative galaxy, and use the color of each segment along the evolution path to indicate the channel that contributes the most to the growth of M_{BH} during a time step. For comparison, we also show the observational M_{BH} versus M_* for both late-type galaxies (LTGs) and early-type galaxies (ETGs) at $z \approx 0$ compiled by Greene et al. (2020). By following the trajectory of each modeled galaxy in the $M_{\text{BH}}-M_*$ plane and the contributions from different channels, One can trace the formation of different types of galaxies and physical processes that drive the growth.

To demonstrate the growth of galaxy ecosystems in further detail, Fig. 14 shows the growth of each mass component, $X = \text{halo, BH, stars or gas}$, contributed by different channels, for three of the four galaxies with colored trajectories in Fig. 13. For each component X , we show the cumulative growth in mass (M_X) and the specific growth rate $\dot{M}_X/M_X/H(z)$. So quantified, the area under the curve of a specific growth rate is the cumulative growth of the mass component

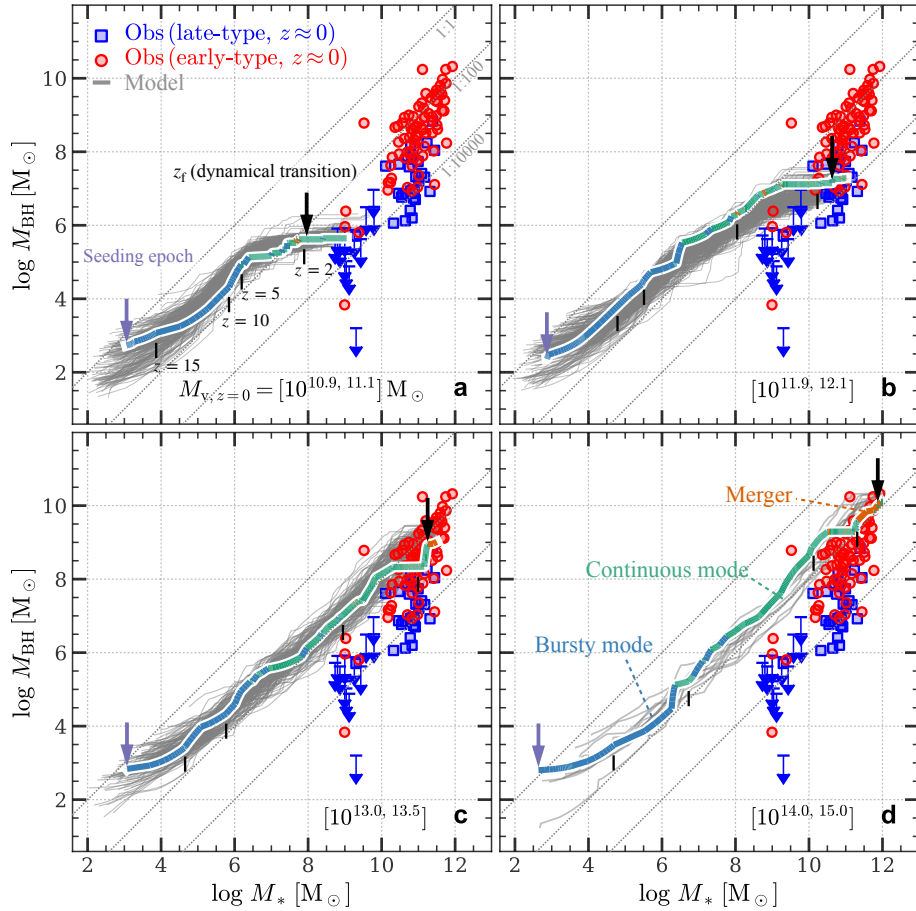


Figure 13. Growth paths of modeled galaxies in the $M_{\text{BH}}-M_*$ plane. a–d, the results for central galaxies with different halo masses at $z = 0$, as labeled. In each panel, **grey** curves show individual modeled galaxies, and a **colored** curve shows one example galaxy with the channel dominating the growth of the BH at each time step indicated by the color of the path segment (**blue**, bursty mode; **green**, continuous mode; **brown**, merger). Note that the dry channel in the bursty mode is always subdominant compared with the wet channel, and is thus totally absent. Critical epochs for the example galaxy are marked (purple arrow: the seeding epoch; black arrow: z_f ; black ticks, $z = 15, 10, 5$ and 2). For reference, **blue** and **red** scatters represent individual late-type and early-type galaxies, respectively, at $z \approx 0$, compiled by Greene et al. (2020). Tilted grey lines indicate the loci of $M_{\text{BH}}:M_* = 1:1, 1:100$ and $1:10000$, respectively. There are too many modeled galaxies in the ranges of $M_{v,z=0}$ in a and b, and we randomly select a subsample of 256 galaxies to display. The channels that dominate the growth of a BH emerge sequentially in time, and are expected to give different observational signatures of the galaxy ecosystem in different ranges of redshift. See §5.3.1 for details.

in logarithmic space. The results of M_X and the specific rate are shown in total as well as separately for the contribution from each channel.

For the dwarf-size galaxies ($M_{v,z=0} \approx 10^{11} M_\odot$), the seeding epoch (marked by a purple arrow in Fig. 13a for the example galaxy) is between $z = 20$ and 30 , consistent with the duration where the cosmic seeding activity peaks (see Fig. 10a). The BH seed has a mass of $M_{\text{BH,seed}} \approx 10^{3.5} M_\odot$ and is certainly a DCBH from a massive progenitor Pop-III star (see Fig. 9 for the mass functions). The birth-time locations of dwarfs in the $M_{\text{BH}}-M_*$ plane is close to the line of $M_{\text{BH}}:M_* = 1:1$, due to the top-heavy IMF assumed for the Pop-III stars that breed the BH seeds (Eq. 34). Such an over-massive BH (relative to the M_* expected from the relation at $z \approx 0$) produces efficient AGN feedback that, together with the efficient stellar feedback in a low-mass halo, prevents the accumulation of star-forming gas into the SGC. Consequently, the mini-galaxy, consisting of a BH and Pop-III stars, becomes quiescent, as seen from horizontal growth paths of M_{BH} and M_* with time (Fig. 14a3 and a5) and the negligible specific growth rates in all channels (Fig. 14a4 and a6) at $z \gtrsim 20$. The growth in the continuous mode remains at low

specific rates of $\lesssim 0.1$ at $z \gtrsim 10$, contributing negligibly to the growth of the BH and stellar masses. However, the specific growth rate of the gas mass within the SGC remains high ($\gtrsim 10$) at these redshifts (Fig. 14a8) due to the strong gas accretion associated with the fast halo assembly, which suggests that the strong AGN feedback is essential to keep the accreted gas hot and to prevent it from forming stars.

This post-seeding quiescent phase, however, relies on the detailed balance between the AGN feedback and the gas inflow, and is thus expected to be highly unstable. A global disturbance of the SGC may drive significant dissipation and temporarily boost the gas inflow to replenish the galactic nucleus. At $z \gtrsim 10$, the frequent excursions of the halo assembly to high γ_v (Fig. 14a2) provide such disturbances, and lead to the episodic formation of the SNF nucleus, that is followed by intense star formation and the super-Eddington accretion of the BH within the nucleus. The consequence of such a bursty growth is the rapid rises of both the BH and stellar masses, as seen from the sharp jumps of M_{BH} and M_* (Fig. 14a3 and a5) and the spikes in the specific growth rates (Fig. 14a4 and a6) at $z \approx 15$. The blue segments in Fig. 13a show that the growth of the dwarf galaxy is dominated by

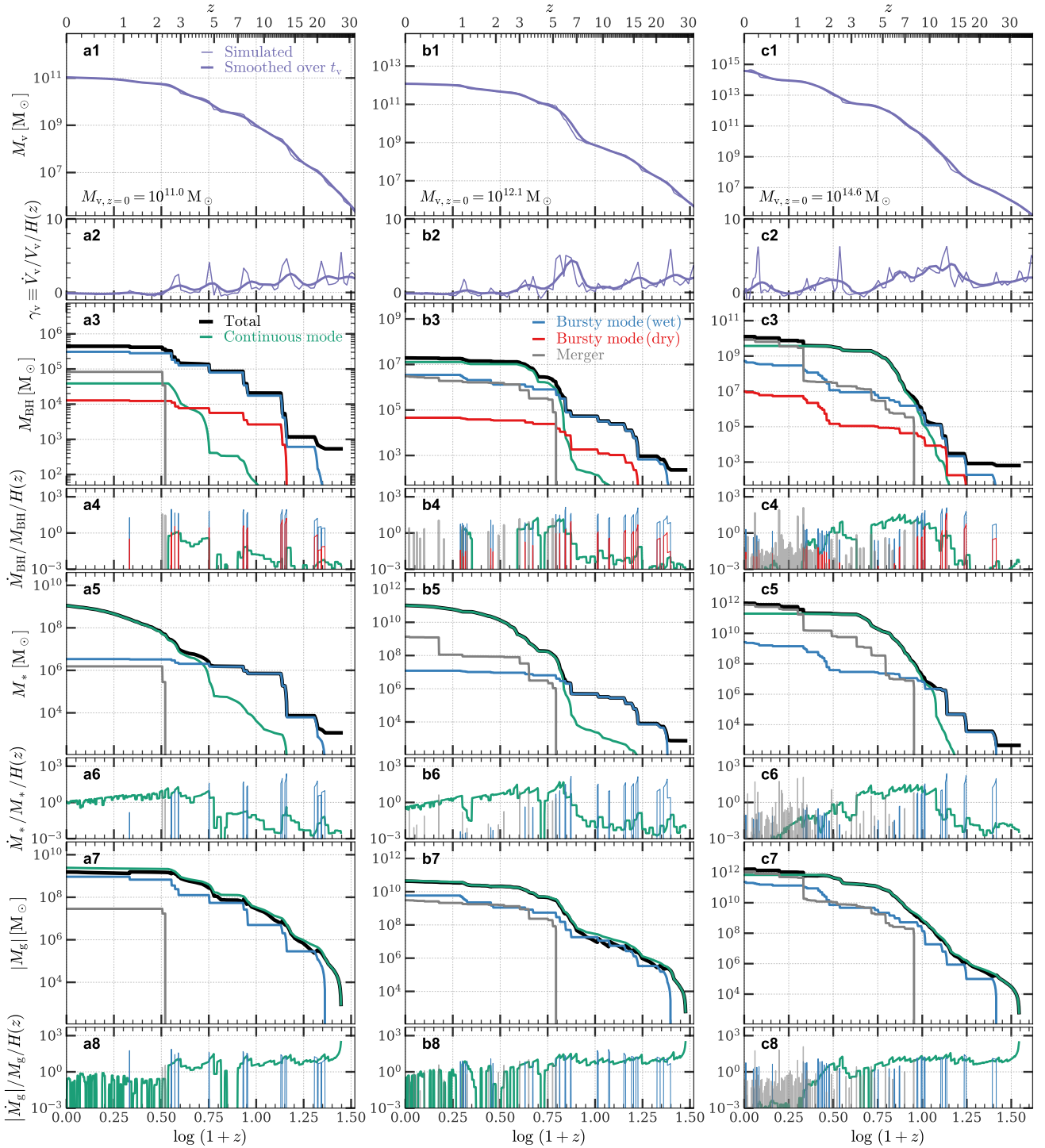


Figure 14. Evolution of different mass components with time and the contributions from different channels. Columns a–c, the results of three example central galaxies (the same galaxies as the colored ones in Fig. 13a, b and d), respectively, with their $M_{v,z=0}$ indicated in the top row. For each component of mass, we show the cumulative growth and specific growth rate in two rows, respectively. (rows 1, 2 for halo mass, M_v ; rows 3, 4 for BH mass, M_{BH} ; rows 5, 6 for stellar mass, M_* ; rows 7, 8 for gas mass, M_g). For halo mass, the simulated growth history and that smoothed by the operator, $\mathcal{S}(\cdot; t_v)$ (Eq. 42) are shown by thin and solid purple curves, respectively. Curves with different colors represent the growth contributed by different channels. Here the channels are: (i) the continuous mode (green), through which star formation extends over the SGC and BH growth moderately fluctuates; (ii) the wet channel in the bursty mode (blue), through which central star formation rises sharply and accretion of gas by BH is boosted until feedbacks turn the nucleus from wet to dry and end the bursty window. For BH, an additional dry channel (red) due to dynamical evolution of NSC is expected to follow the wet stage; (iii) the merger (grey). The total growth, obtained by summing the contributions from all the channels, is shown in black. The contribution of nuclear burst to M_g is negative due to the consumption of gas, so we show the absolute value. The growth rates by nuclear burst (both wet and dry) and merger are evaluated by averaging over 10 Myr, as they fluctuates violently during the bursty window (see Fig. 7 and §4.5). See §5.3.1 for the details.

the wet channel in the bursty mode at $z \gtrsim 5$, and the ratio, M_{BH}/M_* , is kept at $\gtrsim 1/10$. By the time of $z = 5$, the galaxy has reached a stellar mass of $M_* \approx 10^6 M_\odot$ and a BH mass of $M_{\text{BH}} \approx 10^5 M_\odot$, with the growth dominated by the wet channel in the bursty mode (Fig. 14a3 and a5). In contrast, the dry channel of BH growth in the bursty mode is negligible compared to the wet one, because only a small fraction of stars in the nuclear star cluster are captured by the BH in our model implementation (see Appendix A for detail). Thus, the evolution during the bursty stage has two distinct features:

- (i) *the back-and-forth switch between star-forming and quiescent states;*
- (ii) *the formation of a compact, globular-like object expected from the dominance of the nuclear star formation within the SGC.*

Our model thus predicts that *the formation of the first generation of quiescent galaxies is as early as $z \approx 20$ – 30 soon after the seeds are bred, that the population of quiescent galaxies is ubiquitous at $z \gtrsim 10$ as they fall into gaps between nuclear bursts, and that galaxies with compact morphology and prominent nuclear star clusters are common in that era.*

At $z \lesssim 5$, the bursty mode diminishes in the dwarf galaxy due to the less frequent excursion of the halo assembly and the large BH mass compared to the mass of the nucleus (see Fig. 7 for the effects of BH mass). There are still some nuclear bursts at $z = 2$ – 3 when the halo makes significant excursions. Each of the bursts grows the BH mass by about $10^5 M_\odot$ and the stellar mass by about $10^6 M_\odot$. The excursions are caused by halo-halo mergers that also bring in satellite galaxies to merge with the central galaxy later. Mergers in total add about $10^5 M_\odot$ to the BH mass and $10^6 M_\odot$ to the stellar mass (Fig. 14a3 and a5). Due to the low frequency, mergers do not dominate the total mass budgets of BH and stars in the dwarf. The last nuclear burst occurs at $z \approx 1.4$ when the halo assembly fluctuates slightly, causing a negligible growth of M_{BH} and M_* (Fig. 14a3 and a5). In contrast, the continuous mode becomes more significant for star formation, and dominates the cumulative M_* at $z \lesssim 5$ (Fig. 14a5). The specific growth rate of stellar mass catches up with that of the gas mass and falls into the range of $\gtrsim 1$ (Fig. 14a6 and a8), indicating that the continuous mode can sustain a star formation rate comparable to that of the main sequence of star-forming galaxies. The galaxy in this stage mainly builds up its dynamically hot stellar component (bulge), as halo assembly is still in the fast phase ($\gamma_v \gtrsim \gamma_f$; see the beginning of §4.1). The specific growth rate of the BH mass is still less than 1 at $z \lesssim 5$ (Fig. 14a4), and the cumulative M_{BH} contributed by the continuous mode never exceeds that formed in the bursty mode (Fig. 14a3), both reflecting the regulation of the BH growth by AGN and stellar feedback that are efficient in a low-mass halo. At $z \approx 2$, the ratio M_{BH}/M_* is reduced to about $1/100$, a value comparable to massive early type galaxies (ETGs) at $z \approx 0$ (Fig. 13a).

At $z \approx 2$, the host halo of the dwarf galaxy makes a transition to slow assembly, and never recovers to the fast phase again (Fig. 14a1). The SGC thus transits to a dynamically cold disk, with the star formation sustained in the continuous mode at a specific rate of about 1 (Fig. 14a6). In contrast, the BH growth is almost completely quenched due to the slow inward gas migration driven by inefficient viscosity in the cold disk (see the discussion in §4.5.3). In this final stage, a horizontal migration of the galaxy is seen in the $M_{\text{BH}}-M_*$ plane (Fig. 13a), eventually producing a stellar system with $M_* \approx 10^9 M_\odot$ and an under-massive BH with $M_{\text{BH}} \approx 10^{5.5} M_\odot$. Interestingly, the ending points of dwarf galaxies in the $M_{\text{BH}}-M_*$ plane (grey in Fig. 13a) intersect with the upper bounds of observation results compiled by Greene et al. (2020). Since nuclear bursts contribute to the majority of the BH growth in dwarf galaxies, precise measurements of the masses of intermediate-mass black holes

(IMBHs) in dwarf galaxies may put stringent constraints on the physical processes involved in the bursts.

Milky Way (MW)-size galaxies ($M_{v,z=0} \approx 10^{12} M_\odot$) have a growth pattern similar to that of the dwarf galaxies, but with systematic shifts in the values of physical properties. A MW-size galaxy has a seeding epoch that is expected to be slightly earlier than that of a dwarf galaxy due to the earlier formation of the host halo, and has less chance to be bred with a Pop-II seed due to the less intersection with the Pop-II regime in the seeding atlas (Fig. 11a). As in dwarfs, the early growth of the stellar mass and BH mass in an MW-size galaxy is also dominated by the wet channel in the bursty mode (Fig. 13b). Star formation in the continuous mode starts to dominate at $z \approx 5$ – 10 when the M_* reaches about $10^6 M_\odot$ (Fig. 14b5), slightly earlier than that in the dwarfs. A major difference is seen in the growth of the BH mass: the continuous mode of formation in the MW-size galaxy significantly surpasses the M_{BH} accumulated in the bursty mode at $z \lesssim 5$ when M_{BH} exceeds about $10^5 M_\odot$ (Fig. 14b3). This difference can be attributed to the stronger resistance to feedback effects by the deeper potential well of the host halo of an MW-size galaxy – a conclusion also reached by the hydro simulations (Habouzit et al. 2021; Li et al. 2024; Bower et al. 2017) and analytical estimations (Bower et al. 2017; Hong et al. 2023; Paper-I).

At $z \lesssim 3$, the accretion of the host halo of the MW-size galaxy slows down, and eventually stops at $z \approx 1$. During this slow phase, the SGC transits to a dynamically cold disk, and the BH growth is quenched. In contrast, the star formation in the continuous mode reaches the peak efficiency among all galaxies – a discovery first made by observations (e.g. Yang et al. 2003, 2012; Behroozi et al. 2013) and later interpreted as a result of the inefficient stellar and AGN feedback at the MW mass (Croton et al. 2006; Bower et al. 2006; Somerville et al. 2008b). The net effect of the quenched BH growth and efficient star formation in MW-size galaxies is thus a horizontal migration in the $M_{\text{BH}}-M_*$ plane that is most significant among all galaxies (Fig. 13b). The ending points of the migration are at $M_* \approx 10^{11} M_\odot$ and $M_{\text{BH}} \approx 10^7 M_\odot$, with the ratio M_{BH}/M_* close to $1/10000$, which is the lowest among all galaxies. The migration also shapes a $M_{\text{BH}}-M_*$ relation for late-type galaxies (LTGs) that is systematically shifted to the right compared to that for ETGs, consistent with observational results (blue and red dots in Fig. 13b). Similar to the dwarfs, mergers in MW-size galaxies become increasingly important at late time, but do not dominate the mass budgets of BH and stars. In the example shown in Fig. 14b, frequent mergers at $z \lesssim 1$ bring in about $10^9 M_\odot$ of stellar mass, more than that formed in the bursty mode, and about $10^{6.5} M_\odot$ of BH mass, comparable to that grown in the bursty mode.

The brightest cluster galaxies (BCGs; $M_{v,z=0} \gtrsim 10^{13} M_\odot$) follow the general growth pattern found in dwarf and MW-size galaxies, but again with systematical shifts in values of physical properties. However, a distinct, merger-driven evolution pattern appears in their late stages. A BCG is expected to be seeded even earlier, and its BH seed is almost certainly bred in pristine gas environment (Fig. 11b). The transition of the dominant growth mode from the bursty to the continuous occurs at $z \approx 10$, with $M_{\text{BH}} \approx 10^5 M_\odot$ and $M_* \approx 10^6 M_\odot$ (Fig. 14c3 and c5). For the massive example of BCG shown in Fig. 13d and Fig. 14c, the host halo remains in the fast phase until $z = 0$, and the growth of the BH in the continuous mode leads to a BH with a mass of $M_{\text{BH}} \gtrsim 10^{9.5} M_\odot$ at $z = 2$ when the halo reaches a mass of $M_v \approx 10^{13} M_\odot$. After that, the efficient virialization shock and AGN feedback reduce the amount of gas that can cool down into the SGC (Fig. 14c8), and both star formation and BH growth are quenched (Fig. 14c4 and c6). The growth of both BH and stellar masses of the BCG is dominated by the mergers at $z \lesssim 1$, and eventually reaches $M_{\text{BH}} \approx 10^{10} M_\odot$ and $M_* \approx 10^{12} M_\odot$ at $z = 0$,

respectively. A similar conclusion was reached by Li et al. (2024, see their figs. 2 and 10) using a cosmological hydro simulation, who found that the ex-situ growth of SMBHs by merger outpaces the growth by wet accretion in the most massive halos. Other studies using hydro simulations (e.g. Weinberger et al. 2018; Habouzit et al. 2021; Ma et al. 2025a) also highlighted the critical role of mergers in growing the stellar and BH masses in massive galaxies.

One interesting conclusion is that BCGs are the only type of (central) galaxies where all the channels of BH growth are relevant sequentially in the history and become dominant in some stages (as shown clearly by the colored line segments in Fig. 13d). Dynamical hotness in each of the stages plays a fundamental role in maintaining the growth of the BH: in the seeding stage, dynamical heating from the fast assembly of halo delays the collapse of the SGC and allows the formation of massive BH seeds; in the early stage of growth, a temporary excursion of the halo assembly to high γ_v induces global disturbance that leads to the formation of an SNF nucleus and triggers super-Eddington accretion of the BH; when the growth is dominated by the continuous mode, the generation of ballistically moving sub-clouds in the dynamically hot SGC provides the gas supply that sustain the wet growth of the BH; in the late stage when gas cooling is suppressed by the virialization shock and AGN feedback in massive halos, the large number of satellite galaxies that have been brought in by the fast assembly of the host halo provide the central BH with a large number of dry mergers. The sequential dominance of these channels in the history of a BCG yields a high but constant ratio of M_{BH}/M_* , so that the growth path of a BCG in the $M_{\text{BH}}-M_*$ plane almost always stays close to the line $M_{\text{BH}}/M_* = 1/100$ (Fig. 13c and d).

As a brief summary, Figs. 13 and 14 show that the dominant channels sequentially emerging in the growth history of a central galaxy are the nuclear burst, the continuous growth within the self-gravitating gas cloud/disk (with a bursty-to-continuous transition mostly at $z \approx 5-15$), and the ex-situ growth via merger (most significant in BCGs). Depending on the mass and assembly rate of a halo, some of the later channels may be absent in the growth of the BH or stellar mass due to, e.g. dynamical transition to a cold disk and infrequent merger in a low-mass halo.

5.3.2 The cosmic histories of black-hole accretion and star formation

The sequential appearances of the dominant channels in the histories of individual galaxies imply that the cosmic BH accretion history and star formation history are also sequentially dominated by these channels. This conclusion can be understood by considering the cosmic history (of BH accretion or star formation) as a convolution of the halo mass function at $z = 0$ and the histories of individual galaxies with different $M_{v,z=0}$ (e.g. those shown in Fig. 14), plus some contributions from satellite galaxies. The transition epoch of the cosmic history from one channel to the next is thus a weighted average of the transition epochs of galaxies with different $M_{v,z=0}$. Based on the extensive discussion in §5.3.1, we expect that the formation history of galaxies can be divided into the follow eras based on the dominant channel that drives the BH accretion and star formation of galaxies in this era:

- (i) **The seeding era** ($z \gtrsim 20$, during which most seeds have been bred; see §5.2.2 and Fig. 10), in which the cosmic star formation and BH accretion are dominated by the formation of the first generation of stars (mostly Pop-III) and the breeding of BH seeds in mini-halos. Galaxies in this era host over-massive BHs with $M_{\text{BH}}/M_* \approx 1$ (see also §5.4 and Fig. 17i). The latest breeding events of BH seeds in our

sample of halos (within a volume of about 10^6 Mpc^3 , with a minimal halo mass of $M_v \approx 10^{9.5} M_\odot$; see §2) are at $z \approx 5$, in the progenitors of dwarf-size galaxies (see the seeding atlas, Fig. 11). The products of galaxy formation in this era constitute only a small fraction of the total BH and stellar masses observed in the local Universe, given that the mass of BH seeds and the stellar masses of Pop-III star clusters are relatively small compared to those of the typical galaxies at $z \approx 0$. The searching for such tiny fractions of masses at low z is thus challenging.

- (ii) **The bursty era** ($5 \lesssim z \lesssim 20$, before the transition of dwarf-size galaxies to the continuous mode; see §5.3.1 and Figs. 13 and 14), in which BH accretion and star formation take place mostly as episodic nuclear bursts. Galaxies in this era can host a significant fraction of the total stellar mass in nuclear star clusters featured by compact and concentrated morphology. Galaxies can host over-massive BHs, with M_{BH}/M_* kept at around 1/10 (e.g. Fig. 13), until they transit to the continuous mode between $z \approx 5-15$. Quiescent galaxies are expected to be ubiquitous between the gaps of nuclear bursts (Fig. 14, row 5).
- (iii) **The continuous era** ($z \lesssim 5$), in which BH accretion and star formation are both dominated by the continuous mode. Galaxies in this era first appear as bulges and may later transit to disks, with stellar distribution much more extended compared to the nuclear star clusters. BHs are less likely to experience super-Eddington accretion in this era due to the less frequent excursions of the host halos to high γ_v , and the ratios of M_{BH}/M_* are reduced to about 1/100 for galaxies in the dynamically hot (bulge) phase, or even down to about 1/10000 for those in the dynamically cold (disk) phase.

Fig. 15 shows the evolution of the cosmic BH accretion rate (BHAR) density, $\dot{\rho}_{\text{BH}}$, and the star formation rate (SFR) density, $\dot{\rho}_*$, defined respectively as

$$\dot{\rho}_{\text{BH}} = \frac{1}{V_u} \sum_i \dot{M}_{\text{BH},i}, \quad (105)$$

$$\dot{\rho}_* = \frac{1}{V_u} \sum_i \dot{M}_{*,i} / (1 - R), \quad (106)$$

where $R = 0.4$ is the returned fraction of mass by the Chabrier IMF (Bruzual & Charlot 2003) assumed in our model for Pop-I/II stars, and the summations are over all galaxies predicted by our model in the simulated volume, V_u , in each given snapshot. The contributions from BH seeds and Pop-III stars are also included in the total $\dot{\rho}_{\text{BH}}$ and $\dot{\rho}_*$, respectively. The contributions from different channels are also shown in total (black) and separately (colored). For reference, we also show the growth rate of the ‘virialized’ baryon mass density, defined as

$$\text{Virialized } \dot{\rho}_b \equiv \frac{f_b}{V_u} \frac{d}{dt} \sum_i M_{v,i}, \quad (107)$$

scaled down by factors of 10^{-3} and 10^{-2} in the two panels of Fig. 15, respectively, and the summation is over all halos in our sample at each given snapshot. For comparison, we also show results obtained both directly from observations and indirectly from empirical modeling, with (black) or without (grey) JWST.

As expected from the cosmic seeding history (§5.2.2 and Fig. 10), the seeding activity peaks at $z \approx 20-30$, which produces the bumps that mark the beginning of the cosmic BH growth (purple in Fig. 15a). Soon after the seeding, the cosmic BHAR and SFR are dominated by the wet channel in the bursty mode. The contribution of the bursty mode to the cosmic $\dot{\rho}_{\text{BH}}$ rises rapidly with time at $z \gtrsim 10$. This is expected since (i) almost all SGCs have densities close to the supernova-free threshold across their individual bodies (Eq. 10), and thus can each form an SNF nucleus containing a large fraction

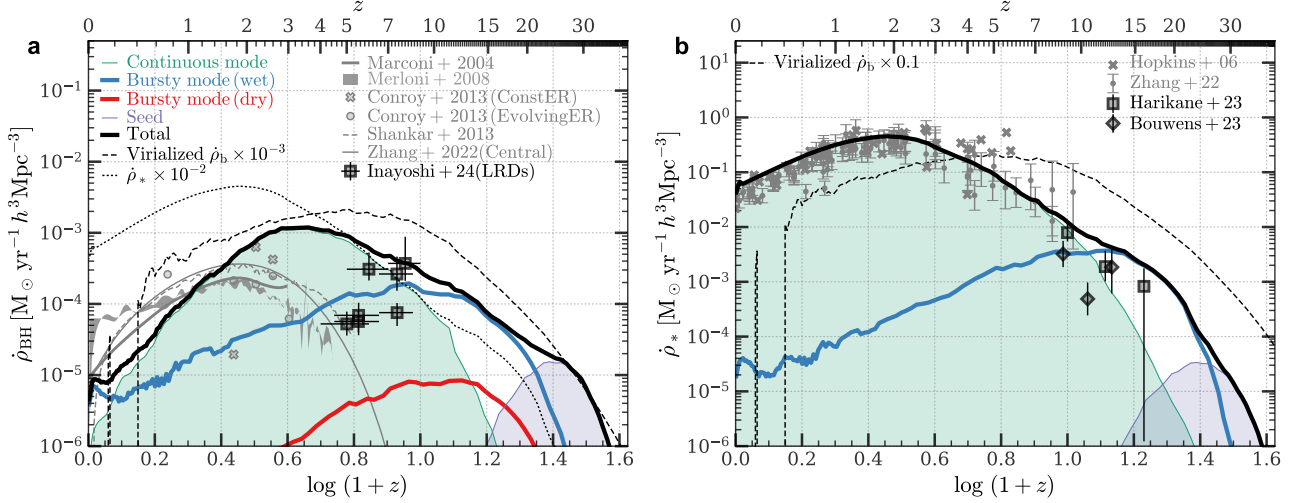


Figure 15. Evolution of cosmic black hole accretion rate (BHAR) density and star formation rate (SFR) density. **a**, cosmic BHAR density ($\dot{\rho}_{\text{BH}}$). **b**, cosmic SFR density ($\dot{\rho}_*$). Both densities are defined as the mass formed per unit comoving volume per unit cosmic time. In each panel, **black** thick curve represents the total rate predicted by our model, while colored curves show the contributions from different channels (**green**, continuous mode; **blue**, wet channel in bursty mode; **red**, dry channel in bursty mode). The contribution from the seeding procedure is also included and shown by **purple** curve. For comparison, we show a fixed fraction of the growth rate of the virialized baryon mass density (virialized $\dot{\rho}_b$), defined as the time derivative of $\sum f_b M_v / V_v$, where the summation is over all halos in the simulated volume. We also scale $\dot{\rho}_*$ by a factor of 10^{-2} and show it in **a**. For $\dot{\rho}_{\text{BH}}$, we show the empirical results calibrated by observations taken from a number of studies, including Marconi et al. (2004), Merloni & Heinz (2008), Conroy & White (2013, with either constant or redshift-dependent Eddington ratio distribution assumed), Shankar et al. (2013) and Zhang et al. (2022, including only central BHs), and the results compiled by Inayoshi & Ichikawa (2024) based on recent JWST observations for the ‘little red dots’ (LRDs; Akins et al. 2024; Kokorev et al. 2024; Matthee et al. 2024; Greene et al. 2024). For $\dot{\rho}_*$, we show the observational data points compiled by Hopkins & Beacom (2006) and Zhang et al. (2022), and recent observational results obtained by Harikane et al. (2023a) and Bouwens et al. (2023) based on JWST. This figure shows that the cosmic histories of BH accretion and star formation can be divided into three sequential eras: the seeding era ($z \geq 20$), the bursty era ($5 \lesssim z \lesssim 20$) and the continuous era ($z \lesssim 5$), with each era dominated by a distinct channel of growth. See §5.3.2 for details.

of the gas mass of the SGC; (ii) halos grow rapidly with time at such high redshift ($\gamma_v \geq 1$ for almost all halos with $M_{v,z=0} \geq 10^{9.5} M_\odot$; see Fig. 3), which provides large amounts of gas fuel for the SGCs; (iii) the SGCs are strongly and frequently disturbed by the frequent excursions of γ_v associated with fast assembly of halos. The growth via the bursty mode during this era efficiently converts about 1/10,000 of the total virialized baryon mass into the BH mass, as seen by a comparison between the dashed and solid black curves in Fig. 15a. However, all three factors that promote the bursty growth diminish with time. At $z \approx 10$ –5, $\dot{\rho}_{\text{BH}}$ of the bursty mode reaches a plateau at the level of about $10^{-4} M_\odot \text{yr}^{-1} h^3 \text{Mpc}^{-3}$ (roughly $10^2 M_\odot \text{yr}^{-1}$ in the entire simulation box), and then declines to $\lesssim 1/10$ of the total $\dot{\rho}_{\text{BH}}$. On the other hand, the contribution of $\dot{\rho}_{\text{BH}}$ from the continuous mode continues to rise, and surpasses that from the bursty mode at $z \approx 7$ –10. This marks the transition from the bursty to continuous era. At $z \approx 3$ –4, $\dot{\rho}_{\text{BH}}$ of the continuous mode reaches a peak at about $10^{-3} M_\odot \text{yr}^{-1} h^3 \text{Mpc}^{-3}$. This is followed by a rapid decline at $z \lesssim 2$ as most halos have now transitioned to the slow assembly phase (see Fig. 1b) in which the BH growth is quenched. The cosmic $\dot{\rho}_{\text{BH}}$ during this era drops to a level that is significantly below 1% of $\dot{\rho}_*$, consistent with the expectation from the low values of M_{BH}/M_* for the LTGs at $z \approx 0$ (Fig. 13). At $z \approx 0$, the bursty mode appears outpacing the continuous mode again, but the rate of nuclear bursts and the growth of BHs during each burst are small, as seen from the large fluctuations in, and small values of $\dot{\rho}_{\text{BH}}$.

The cosmic SFR density $\dot{\rho}_*$ (Fig. 15b) has an evolution pattern similar to that of $\dot{\rho}_{\text{BH}}$. The contribution of the bursty mode to $\dot{\rho}_*$ dominates the era after the peak of the seeding activity, and rises rapidly with time at $z \geq 15$. Bursty star formation in this era converts about 1% of the virialized baryon mass into stars, comparable to

that in dwarf galaxies but much lower than that in MW-size galaxies in the local Universe. This indicates an overall low-efficiency of star formation through the bursty mode, as the depletion of the nuclear gas is mainly by the AGN-driven outflow rather than the consumption from star formation (see Eq. 4.5.1 for the conclusions based on depletion timescales). The epoch that the continuous mode starts to surpass the bursty mode in $\dot{\rho}_*$ is at $z \approx 10$, slightly earlier than that in $\dot{\rho}_{\text{BH}}$, due to the competition between BH accretion and star formation in the growth of the BH and star formation in the continuous mode. At $z < 5$, the contribution of the bursty mode makes up only $\lesssim 10\%$ of the total $\dot{\rho}_*$, indicating that the formation has transitioned to the continuous era. The subsequent evolution of the continuous-mode star formation follows the cosmic $\dot{\rho}_*$ history obtained in observations (grey markers), with a single peak at $z \approx 2$, the ‘cosmic noon’ of star formation. The virialized $\dot{\rho}_b$ shows a decline in this era, as expected from the slowdown of the halo assembly. Star formation around and after the cosmic noon can convert a large fraction of the virialized $\dot{\rho}_b$ into stars, indicating that some of the baryons ejected by previous feedback processes can be recycled to the galaxies to fuel the star formation.

Recent JWST observations have extended the constraints for the cosmic histories of BH accretion and star formation to redshifts higher than previously possible. Remarkably, a population of ‘little red dots’ (LRDs), featured by their peculiar V-shaped spectral energy distributions (SEDs), extremely compact morphologies and large population density (about 100 \times that of UV-selected quasars) have been observed at $z \geq 4$ (Akins et al. 2024; Barro et al. 2024; Greene et al. 2024; Kokorev et al. 2024; Matthee et al. 2024; Pacucci & Narayan 2024; Yue et al. 2024b; Chen et al. 2025b; Labbe et al. 2025;

Li et al. 2025c; Perger et al. 2025). Analyses based on the compact morphologies and the presence of broad emission lines suggest that LRDs may harbor over-massive BHs with M_{BH} comparable to or higher than a few percent of the M_* of the host galaxies (Chen et al. 2025b). The high BH masses and high abundance of LRDs, if confirmed, will significantly increase the cosmic BHAR density. In Fig. 15a, we show the compilation by Inayoshi & Ichikawa (2024, see their fig. 2) based on a number of LRD samples at $z \gtrsim 4$. The values of the cosmic $\dot{\rho}_{\text{BH}}$ obtained are much higher than the extrapolation from the empirical results at $z \lesssim 4$ obtained prior to JWST (grey curves and scatters; Marconi et al. 2004; Merloni & Heinz 2008; Conroy & White 2013; Shankar et al. 2013; Zhang et al. 2022). Interestingly, the contribution of the bursty mode to BH accretion in our model matches the range of the LRD-based observations in the $\dot{\rho}_{\text{BH}}-z$ plane. Our model thus has the potential to explain the LRD population as the mixture of nuclear star clusters and over-massive BHs formed in nuclear bursts.

The JWST observations of the cosmic $\dot{\rho}_*$ also suggest a moderate enhancement of the star formation efficiency in the early Universe (Naidu et al. 2022; Chen et al. 2023b; Yung et al. 2023; Harikane et al. 2023a; Xiao et al. 2024; Sabti et al. 2024; Wang et al. 2025). The predictions of our model at $z \gtrsim 10$ shown in Fig. 15b are consistent with the JWST results, even though the model is not calibrated to match these observations. Since star formation at such high redshift is expected to be dominated by the bursty mode, the precise measurements of the cosmic $\dot{\rho}_*$ at $z \gtrsim 10$ may offer clean constraints on processes affecting the properties of nuclear starbursts.

5.3.3 Effects of seeding methods

The small contributions of BH seeds and Pop-III star clusters to the cosmic BHAR and SFR densities shown in Fig. 15 imply that these earliest products of galaxy formation are soon mixed into, and outshone by the BH and stellar masses formed after the seeding era. As current instruments are not sensitive enough to observe the active Pop-III star formation and BH seeding at $z \gtrsim 20$, efforts to distinguish seeding scenarios have been made using targets at lower redshift, $z \approx 5-15$ (e.g. D’Antona et al. 2023; Pacucci et al. 2023; Onoue et al. 2024; Jeon et al. 2025), when the BH and stellar masses of galaxies are already significantly modified by the bursty mode in our model. Thus, it is important to understand whether and how the seeds bred by different methods preserve some observable signatures after the modification by subsequent growth. Here we perform an *idealized and controlled experiment* for each target halo by using the two extreme variants of our model, referred to as ‘MinSeed’ and ‘MaxSeed’, to bracket the seed masses expected from variants covering the entire parameter space. All the other ingredients of the model, including the input subhalo merger trees, physical prescriptions governing the post-seeding growth of BHs and stars, and even the random number sequences implemented for stochastic processes in the model, are kept strictly the same. Such a controlled experiment allows one to isolate the effects of seeding methods and to test whether certain observational probes are capable of distinguishing seeds bred by different methods.

For the above purpose, we define the extreme variants within the Default model, as follows.

- (i) We disable the IGM enrichment by forcing the gas around any halo at any redshift to be pristine (i.e. $Z_{\text{IGM}} = 0$). This ensures that Pop-III seeds are bred and the difference of $M_{\text{BH,seed}}$ between the two variants can be maximized without the entanglement introduced by Pop-II seeds (see Fig. 11).
- (ii) For the ‘MinSeed’ variant, we disable the LW background by forcing

$J_{\text{LW},21} = 0$ for any halo, at any redshift, and we set the parameter $\gamma_{\text{crit}} \rightarrow +\infty$. This ensures that the factors f_j (Eq. 18) and f_γ (Eq. 22) are both zero, so that the delay of gas collapse in mini-halos is completely disabled. Since Pop-III clusters form once the mini-halos reach the H_2 -cooling threshold (Eq. 14 and Fig. 1), the resulted $M_{\text{BH,seed}}$ in this case is thus minimal among all the possible variants in the model space. In contrast, for the ‘MaxSeed’ variant, we force $J_{\text{LW},21} \rightarrow +\infty$ and $\gamma_{\text{crit}} \rightarrow 0$, so that the delay of seeding and the masses of seeds are maximized.

To cover a large range of halo masses and to have a sufficiently large sample, here we switch to halo merger trees generated by the EPS-based Monte-Carlo method (see Appendix A1). For a target halo with a given mass at $z = 0$ ($M_{\text{V},z=0}$), we randomly generate a merger tree rooted in the halo. Since we focus on galaxies at high z when the differences between the variants are significant (see below), mergers do not play an important role in the growth of M_{BH} and M_* (see §5.3.1 and Fig. 13). Thus, we prune the side branches and only retain the main branch of each tree to save computational cost. Note that such a pruning only removes the ‘dry’ contributions to M_{BH} and M_* from side branches, i.e. masses already contained in the side branches and added to the main branch during mergers, but the ‘wet’ contributions, i.e. masses added by the continuous and bursty modes due to mergers, are still retained by including the excursions of halo assembly. The two variants of the model are then applied independently to the main branch of the same target halo, and produce two sets of growth histories, respectively, for the central galaxy in the halo that differ only in the seeding methods (plus the post-seeding processes that depend on the results of seeding; see below).

In Fig. 16, we show the predictions obtained from the two variants for example halos, each with a given $M_{\text{V},z=0}$ (colored thin and thick curves, respectively). A sample of 500 halos with the same $M_{\text{V},z=0}$ as each example halo is shown (shaded region, obtained by the MaxSeed variant). Both M_{BH} and M_* are displayed as a function of the cosmic time, and the ratios between the two variants are also shown. Recent JWST observations for the M_{BH} of a small set of BHs and the M_* of their host galaxies are also included for comparison. In the MinSeed variant, BH seeds with masses $M_{\text{BH}} \approx 10^2 M_\odot$ are bred immediately after halos reach the H_2 -cooling threshold (Eq. 14) at $z \gtrsim 20$, during an epoch when the Universe is about 100 Myr old. In the MaxSeed variant, as expected from the seeding atlas (Fig. 11), the breeding of BH seeds is delayed, more so for less massive halos, and the seed masses are raised to $M_{\text{BH}} \gtrsim 10^4 M_\odot$, with DCBH flavors (see the mass functions in Eq. 9). For the example galaxy with $M_{\text{V},z=0} = 10^{10} M_\odot$, the seeding is delayed most significantly, to $z \approx 8-9$, a range of redshifts that can be covered by JWST. The differences in the breeding redshift and $M_{\text{BH,seed}}$ between the two variants highlight the critical role of environmental effects in the seeding process, as detailed in §5.2.3.

The large difference in $M_{\text{BH,seed}}$ implies that the strength of AGN feedback also differs significantly between the two variants. For the growth in the bursty mode, a larger M_{BH} implies a higher Eddington luminosity L_{Edd} , and thus a quicker depletion of the gas in the SNF nucleus (Eq. 88). Such a trend can be seen in the examples given by the refined treatment of nuclear bursts (Fig. 7), where a BH with a larger initial mass ends the burst earlier, and produces smaller jumps in M_{BH} and M_* (in logarithmic scale) than that with a smaller initial mass. For the growth in the continuous mode, a larger M_{BH} causes stronger ejection and prevention that reduce the accumulation rate of star-forming gas in the SGC and the formation rate of sub-clusters in the gas, thus limiting the growth of both M_{BH} and M_* . In our model, this effect is described by the factor F_{AGN} (Eq. A7), which predicts the fraction of star-forming gas after the regulation by AGN feedback.

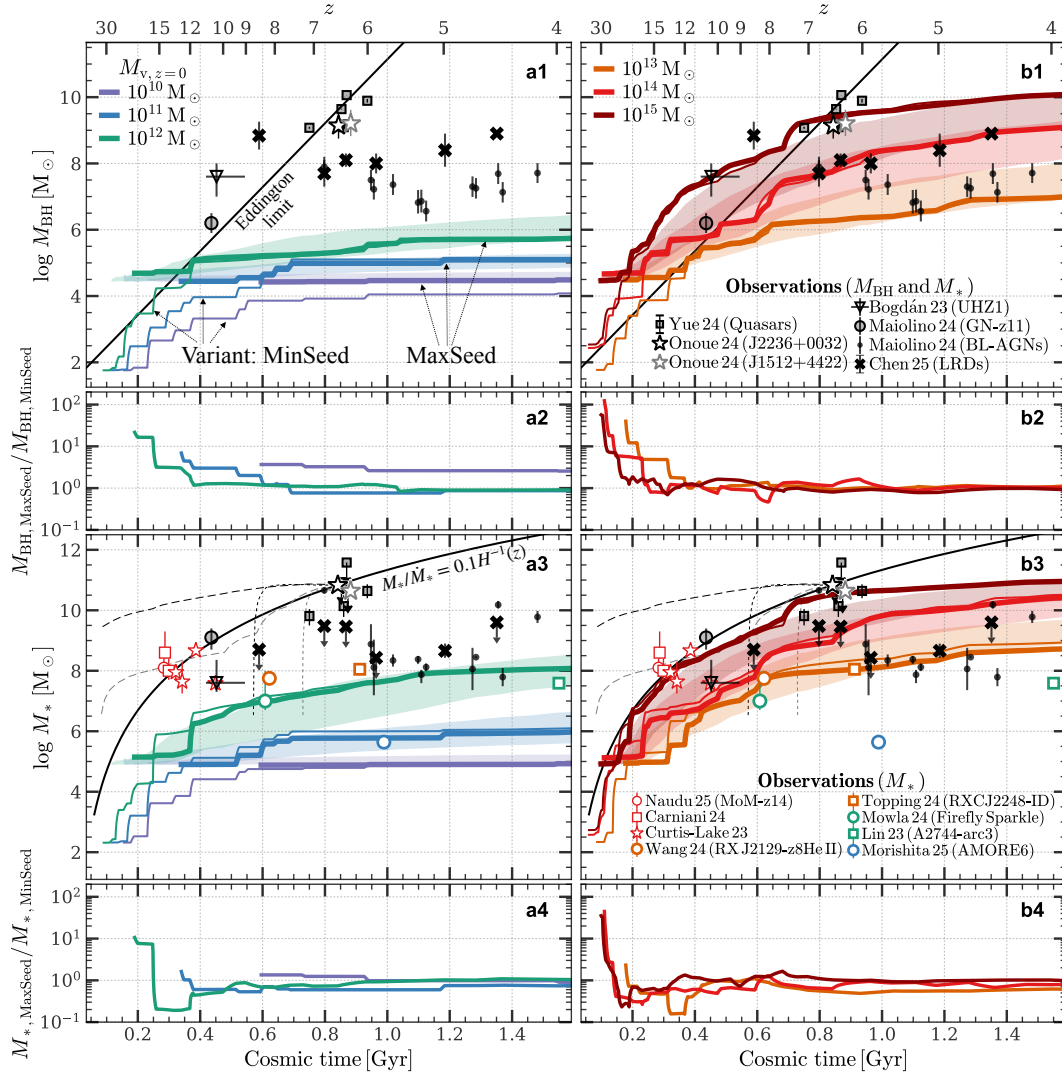


Figure 16. Idealized and controlled experiments showing the effects of seeding method and demonstrating the idea of archaeology and futurology. Each colored curve shows the growth history predicted by our model for the central galaxy in a halo with a given halo mass at $z = 0$. The results are separated into two columns for clarity: **a**, $M_{v,z=0} = 10^{10}, 10^{11}$ and $10^{12} M_{\odot}$; **b**, $M_{v,z=0} = 10^{13}, 10^{14}$ and $10^{15} M_{\odot}$. For each halo, the ‘MinSeed’ and ‘MaxSeed’ variants of our model are applied, and the results are shown by **thin** and **thick** curves, respectively. The two variants differ only in the seeding methods, producing minimal and maximal masses of BH seeds, respectively, among all the variants in the entire model space. All other ingredients of the model are controlled to be strictly the same as the ‘Default’ model, thus isolating the effects of seeding methods on the growth of galaxies. Each row shows a property as a function of cosmic time: **1**, BH mass (M_{BH}); **2**, the ratio of M_{BH} between the ‘MaxSeed’ and ‘MinSeed’ variants; **3**, stellar mass (M_*); **4**, the ratio of M_* between the ‘MaxSeed’ and ‘MinSeed’ variants. To demonstrate the diversity of halo assembly, **shaded** region covers the $1-\sigma$ (16th–84th) percentiles, at each redshift, among the masses predicted by the ‘MaxSeed’ variant for a sample of 500 galaxies with $M_{v,z=0}$ the same as the example halo. In **a1** and **b1**, a tilted **black line** shows the path of M_{BH} expected for a BH growing at the Eddington rate (with a radiative efficiency of $\epsilon_r = 0.1$, i.e. $\dot{M}_{BH} = 10M_{BH}/t_{Sal}$ in our notation; see §4.3.2). In **a3** and **b3**, a **black solid curve** shows the path of M_* expected for a galaxy with a specific star formation rate (sSFR $\equiv \dot{M}_*/M_*$) equal to $t_v^{-1} = 10H(z)$ (see Eq. 7). For comparison, we show the JWST observations of M_{BH} and M_* , obtained by [Yue et al. \(2024a\)](#) for galaxies hosting luminous quasars at $z \geq 6$; [Onoue et al. \(2024\)](#) for two massive ‘post-starburst’ galaxies hosting quasars at $z \approx 6-7$; [Bogdán et al. \(2023\)](#) for the galaxy UHZ1; [Maiolino et al. \(2024a\)](#) for the galaxy GN-z11; [Maiolino et al. \(2024b\)](#) for a sample of broad-line AGNs (BL-AGNs); and [Chen et al. \(2025b\)](#) for a sample of ‘little red dots’ (LRDs). Dashed and solid curves connecting the post-starburst galaxies in **b** represent the star formation histories, assuming non-parametric and delayed- τ forms, respectively, obtained by their spectrophotometric fitting method. In row 3, we also show the JWST observations of M_* , obtained by [Naidu et al. \(2025\)](#) for a Lyman-break galaxy ‘MoM-z14’; [Carniani et al. \(2024\)](#) for two Lyman-break galaxies, ‘JADES-GS-z14-0’ and ‘JADES-GS-z14-1’; [Curtis-Lake et al. \(2023\)](#) for four metal-poor galaxies at $z \approx 10-13$; [Wang et al. \(2024b\)](#) for a HeII emitter at $z \approx 8$; [Topping et al. \(2024\)](#) for a nitrogen-rich galaxy at $z \approx 6$; [Mowla et al. \(2024\)](#) for a galaxy ‘Firefly Sparkle’ at $z \approx 8$ that appears as a set of dense star clusters; [Lin et al. \(2023\)](#) for a galaxy ‘A2744-arc3’ at $z \approx 4$ with dense star clusters and inverted metallicity gradient; [Morishita et al. \(2025\)](#) for ‘AMORE6’, a low-mass, metal-poor galaxy at $z \approx 6$. EPS-based Monte Carlo trees are used as input of the model for this figure to ensure a sufficiently large number of massive halos can be obtained. This figure shows that the BHs produced by different seeding methods converges quickly to a similar M_{BH} owing to the post-seeding growth in the bursty mode. Matching the observed galaxies with predicted evolutionary tracks can be performed to reconstruct the formation history of the galaxies, and to forecast their future growth. See §§5.3.3 and 5.3.4 for the details.

Since this factor is a monotonically decreasing function of M_{BH} , the post-seeding growth of M_{BH} and M_* in both modes is expected to be more limited if the BH seed bred in the mini-halo is more massive. As a result, the difference in BH mass produced by seeding methods may be erased by the post-seeding growth.

The growth histories of M_{BH} produced by our controlled experiments confirm the above expectation. In (the progenitors of) the massive halos with $M_{\text{v},z=0} \gtrsim 10^{13} M_{\odot}$ (Fig. 16b1 and b2), the average growth rates of M_{BH} in the MinSeed variant during the post-seeding period ($z \gtrsim 12$) are around or above the Eddington limit (indicated by a black tilted line in panel b1). This era of rapid BH growth is driven by the bursty mode, as already explained in §5.3.1. Each individual nuclear burst produces a significant growth of M_{BH} via super-Eddington accretion within a short timescale of $\lesssim 1$ Myr, and manifests itself as a vertical jump in the evolution path of M_{BH} in panel b1. In contrast, the growth of M_{BH} in the MaxSeed variant exhibits jumps with much smaller amplitudes due to the stronger AGN feedback. Such a difference in the post-seeding growth of M_{BH} between the two variants reflects the self-regulation nature of BH growth, a conclusion also reached from hydro simulations by analyzing the co-evolution of BHs with their host galaxies (Li et al. 2024) and by controlled experiments (Su et al. 2025; Sivasankaran et al. 2025). The self-regulated growth of BH thus drives the paths of M_{BH} in the two variants to converge to a similar value, and almost completely erases the difference in M_{BH} produced by the seeding methods.

The mass of the BH at the convergence is about $10^5 M_{\odot}$ for all the example halos in Fig. 16b1 and b2. The epoch of the convergence, however, varies with the halo mass, from $z \approx 20$ for the case of $M_{\text{v},z=0} \approx 10^{15} M_{\odot}$ to $z \approx 12$ for $M_{\text{v},z=0} \approx 10^{13} M_{\odot}$ (panel b2). Given the low BH mass and high redshift expected before the convergence and the negligible difference in the BH mass after the convergence, the results of our controlled experiments suggest that it may be observationally challenging to distinguish the seeding methods by measuring M_{BH} .

For less massive halos with $M_{\text{v},z=0} \lesssim 10^{12} M_{\odot}$, the growth of BHs in the controlled experiments (Fig. 16a1 and a2) follow a pattern similar to that in the massive halos. Low-mass seeds bred in the MinSeed variant experience more significant jumps of M_{BH} during individual nuclear bursts than high-mass seeds bred in MaxSeed. However, the average growth rate of the BH in a low-mass halo during the post-seeding era is around or below the Eddington limit due to the less frequent excursions of halo assembly to high γ_{v} and less amount of gas supply in the SNF nuclei (see §5.3.1 and Fig. 14). As a result, the convergence of the growth paths of M_{BH} between the two variants is delayed, at about $z \lesssim 12$ for the case of $M_{\text{v},z=0} \approx 10^{12} M_{\odot}$, and $z \approx 8$ – 9 for $M_{\text{v},z=0} \approx 10^{11} M_{\odot}$. Interestingly, the convergence never occurs for the case with the smallest halo mass, $M_{\text{v},z=0} \approx 10^{10} M_{\odot}$. The progenitors of dwarf galaxies with $M_{\text{v},z=0} \lesssim 10^{11} M_{\odot}$ and the IMBHs hosted by them may thus provide an opportunity to distinguish the seeding scenarios.

The growth of stellar mass in the controlled experiments also follows a similar pattern as that of M_{BH} . The MinSeed variant produces a stellar mass of the Pop-III star cluster that is significantly lower than that in MaxSeed, as expected from the nearly linear relation between $M_{\text{BH,seed}}$ and M_* (Eq. 37) implied by the adopted IMF. The most massive example with $M_{\text{v},z=0} \approx 10^{15} M_{\odot}$ (Fig. 16b3) closely follow the path $\text{sSFR} = \tau_{\text{v}}^{-1}$ (indicated by a black solid curve), while the less massive examples grow with lower sSFR. The convergence of the growth paths of M_* between the two variants are also associated with the self-regulation of the BH, and the convergence is reached earlier in the halo with a higher mass. The converging epochs of

M_* , however, are systematically earlier than those of M_{BH} , owing to the less competitive AGN feedback on the SNF nuclei from BHs of lower masses and the more efficient star formation during nuclear bursts (see §4.5.1 and Fig. 7). Quickly after the seeding, no observable difference in the stellar mass between the two variants can be found, implying that measurements of M_* are not capable of providing useful constraints on the seeding methods.

5.3.4 Archaeology and futurology: linking black holes and galaxies across cosmic time

A useful application of the controlled experiments is to provide a prior for model inferences from observations. Here we demonstrate the idea by recovering the growth histories of observed galaxies through galactic archaeology and model forecasting. In Fig. 16, we over-plot the JWST-based measurements of redshift and masses (row 1, M_{BH} ; row 3, M_*) for individual galaxies. By simply comparing the locations of the observed galaxies in the $M_{\text{BH}}-z$ and M_*-z planes with the growth histories predicted by our model, useful conclusions for the general population of observed galaxies can already be made. All the observations fall into the converged regions of the MinSeed and MaxSeed variants, suggesting that (i) the observed galaxies cannot be used to distinguish the seeding methods, as seen in §5.3.3; (ii) the backward archaeology of the growth histories for the observed galaxies is limited to the range of redshift after the converging epoch ($z \gtrsim 12$ for $M_{\text{v},z=0} \gtrsim 10^{13} M_{\odot}$), with a barrier introduced by the degeneracy between the seeding methods and the self-regulated growth of BH; (iii) the forward forecasting of the growth histories for the observed galaxies is not sensitive to the seeding methods. The growth histories of galaxies in massive halos are very diverse: at fixed $M_{\text{v},z=0} \gtrsim 10^{15} M_{\odot}$ the 1- σ percentile range of M_{BH} (M_*) is about 2 dex (1 dex) at $z \gtrsim 4$. The growth history of a given galaxy can fluctuate significantly, and often exhibits rapid jumps and flat periods, owing to the back-and-forth switch between star-forming and quiescent states (§5.3.1). Such large intrinsic diversity and fluctuation in the growth histories of galaxies imply that there is large chance for observations to catch extreme objects, such as those with M_{BH} and M_* much larger than those expected from a halo mass function with an average formation efficiency (Chen et al. 2023b; Boylan-Kolchin 2023; Maiolino et al. 2024a; Pacucci & Loeb 2024).

By comparing the observations with the growth histories of the modeled galaxies with different $M_{\text{v},z=0}$ (Fig. 16), the archaeological and forecasting information of individual galaxies can be obtained, and the objects observed at different redshifts may be linked together to form a coherent view of the growth history. The two massive, post-starburst quasar-host galaxies at $z \approx 6$ – 7 observed by Onoue et al. (2024) appear to overlap with the evolutionary paths of M_{BH} and M_* of the example BCG with $M_{\text{v},z=0} = 10^{15} M_{\odot}$. Compared with the ensemble of the growth paths of galaxies in such halos (shaded region), the observed M_{BH} and M_* are slightly above the upper 1- σ (84th) percentile, indicating that the two galaxies are massive outliers even among the BCG progenitors of Coma-size halos ($M_{\text{v}} = 10^{14.96} M_{\odot}$; Okabe et al. 2014). The quasar-host galaxies reported by Yue et al. (2024a) at $z \gtrsim 6$ occupy the ranges of M_{BH} and M_* similar to those of the two post-starburst galaxies, suggesting that they have similar origins and formation pathways. Among the sample of LRDs with M_{BH} and M_* reported by Chen et al. (2025b), three (at $z = 4.46$, 6.34 and 6.76) are within the 1- σ ranges of M_{BH} and M_* predicted for BCG progenitors with $M_{\text{v},z=0} \approx 10^{14} M_{\odot}$; two (at $z = 4.96$ and 5.84) have M_* (or upper bound) close to the example BCG progenitor with $M_{\text{v},z=0} \approx 10^{13} M_{\odot}$, but their M_{BH} are over-massive, close to that for $M_{\text{v},z=0} \approx 10^{14} M_{\odot}$; one (at $z = 8.50$) has an obviously

over-massive M_{BH} , even compared with the $1\text{-}\sigma$ range of the BCG progenitors with $M_{\text{V},z=0} \approx 10^{15} M_{\odot}$. The broad-line AGNs reported by Maiolino et al. (2024b) are consistent with the BCG progenitors with $M_{\text{V},z=0} \approx 10^{13} M_{\odot}$, but some have under-massive M_* that fall into the $1\text{-}\sigma$ range of the modeled galaxies with $M_{\text{V},z=0} \approx 10^{12} M_{\odot}$.

Towards earlier epochs, the galaxy GN-z11 at $z = 10.60$ (Maiolino et al. 2024a; see also Tacchella et al. 2023; Ji et al. 2025) has a M_* close to the example BCG with $M_{\text{V},z=0} \approx 10^{15} M_{\odot}$ (Fig. 16, column b), making it a viable progenitor for the two post-starburst galaxies (Onoue et al. 2024) and other quasar-host galaxies (Yue et al. 2024a). Its M_{BH} is significantly under-massive, but still consistent with the $1\text{-}\sigma$ range of the BCG progenitors with $M_{\text{V},z=0} \approx 10^{15} M_{\odot}$. A galaxy hosted by such a high mass halo is predicted to have experienced repeated nuclear bursts by this epoch, and thus to have built a compact and massive nuclear star cluster around the BH (see §5.3.1 and Fig. 14). This is consistent with the observational inference by e.g. D’Antona et al. (2023) that GN-z11 was formed out of pristine gas mixed with ejecta from asymptotic giant branch (AGB) stars in a massive globular or nuclear cluster. A similar conclusion can be reached for the galaxy UHZ1 at $z \approx 10.3$, which hosts a heavily obscured but X-ray luminous quasar (Bogdán et al. 2023), and for the samples of Lyman-break galaxies at $z \gtrsim 10\text{--}14$ (reported by Naidu et al. (2025), Carniani et al. (2024) and Curtis-Lake et al. (2023)) whose M_* align with the evolutionary paths predicted by the model for BCG progenitors with $M_{\text{V},z=0} \approx 10^{15} M_{\odot}$. Thus, their fates are expected to be the bright quasars at $z \gtrsim 6$ and the BCGs of Coma-size halos at $z \approx 0$.

Observations for galaxies with lower M_* in the early Universe are more challenging, but have been achieved by JWST with strong lensing. The cases include those shown in row 3 of Fig. 16, obtained by Mowla et al. (2024, ‘Firely Sparkle’), Lin et al. (2023, ‘A2744-arc3’), and Morishita et al. (2025, ‘AMORE6’). The former two are occupied by compact star clusters, and the latter is a galaxy with a compact morphology. All the three have peculiar enrichment pattern: Firely Sparkle has ISM and stellar metallicities of about 1% of solar values; A2744-arc3 has ISM metallicity that appears to be lower than that of the CGM; AMORE6 has an extremely low ISM metallicity of $\lesssim 0.1\%$ of the solar value. Both the peculiar morphology and the peculiar metallicity of these galaxies suggest that they are in the early stage of evolution, where an extended, smooth stellar body and well-enriched, well-mixed gas environment have not yet been established. The observed M_* of these galaxies are consistent with the evolutionary paths of modeled galaxies with $M_{\text{V},z=0} \approx 10^{11}\text{--}10^{12} M_{\odot}$ (Fig. 16a3). The prediction is that a large fraction of their stellar mass is formed by nuclear bursts or in the form of globular clusters (Chen et al. 2025a, see their figs. 17 and 18), which provides a plausible explanation for their peculiar properties.

Note that the mismatch of individual galaxies between model predictions and observations are not unexpected, given that our model has not been calibrated to match the observations, and that the observations themselves are subject to systematics introduced by, e.g. sample selection, the degeneracy in SED/spectral fitting, and the extrapolation of spectral templates. Examples of such uncertainties are demonstrated in row 3 of Fig. 16 by the star formation histories of the two post-starburst galaxies obtained by Onoue et al. 2024. Their spectrophotometric fitting with two different prior forms for the star formation histories (non-parametric and delayed- τ) yield unconstrained results at high z , highlighting the challenges in observational data modeling. Observational data on galaxy environments may provide additional information about the formation history, as demonstrated by using 2PCCFs for galaxies with different seeds in our model (§5.2.3 and Fig. 12) and for galaxies with different struc-

ture parameters reported in a recent observational study (Zhang et al. 2025a). Similarly, our model can be implemented into a constrained N-body simulation (e.g. Wang et al. 2016; Stiskalek et al. 2025) in which the large-scale dark-matter environments around individual halos are constrained, leaving the modeling of baryonic (Z_{IGM}) and radiative ($J_{\text{LW},21}$) environments for the seeding less uncertain.

5.4 The build-up of the black hole mass-stellar mass relation

The growth pattern of SMBHs and their host galaxies detailed above provide physical insights into how individual galaxies build up their BH and stellar contents. In observation, such evolutionary information of individual objects is often not accessible. Instead, properties of objects can be measured for representative samples within discrete intervals of redshift, and relations between the properties can be constructed to form a statistical description for galaxies at each snapshot of the Universe. Since the imprints of physical processes driving galaxy formation are encoded in these relations, albeit in a degenerate way, retrieving physical information from the relations has been a major endeavor in studies to link observations to formation processes.

One of the best-studied relations describing the co-evolution of SMBHs and their host galaxies is the BH mass-stellar mass ($M_{\text{BH}}\text{-}M_*$) relation. The relation was measured observationally for galaxies in the local Universe (e.g. Kormendy & Ho 2013; Greene et al. 2020; Zhuang & Ho 2023; Graham & Sahu 2023) and has been recently extended to high redshift by observations with e.g. JWST and ALMA (Izumi et al. 2021; Übler et al. 2023; Harikane et al. 2023b; Maiolino et al. 2024b; Onoue et al. 2024; Chen et al. 2025b). Theoretical explanations have also been proposed to link the relation to physical processes that drive the co-evolution of SMBHs with their host galaxies and halos (Bower et al. 2017; Habouzit et al. 2021; Li et al. 2024; Hong et al. 2023; Voit et al. 2024). In our model, the $M_{\text{BH}}\text{-}M_*$ relation is not expected to be a single power-law relation, which can be understood through the following reasoning: (i) The dominant channels (seeding, bursty or continuous mode, or merger) sequentially emerge in the growth histories of individual galaxies (§5.3.1 and Fig. 13). The sub-population of galaxies within a certain stage where the growth is dominated by one channel is thus expected to exhibit a distinct growth path in the $M_{\text{BH}}\text{-}M_*$ plane and to share a relation with distinct properties. (ii) The halo mass ($M_{\text{V},z=0}$) appears to be an important factor to determine the growth stage of a galaxy at a given redshift, as seen from Figs. 13 and 14. Thus, galaxies with similar $M_{\text{V},z=0}$ and z are expected to form a sub-population that share a similar $M_{\text{BH}}\text{-}M_*$ relation. (iii) The overall $M_{\text{BH}}\text{-}M_*$ relation at a given z is a convolution between the halo mass function at $z = 0$ and the relations of the sub-populations at z .

Here we show the $M_{\text{BH}}\text{-}M_*$ relations at different z predicted by the Default variant of our model, and demonstrate how the relations are built up by different formation channels. In Appendix C, we explore the effects of changing model prescriptions on the predicted $M_{\text{BH}}\text{-}M_*$ relation.

5.4.1 Evolution of the relation

Fig. 17 shows the distribution of galaxies (grey dots) predicted by our model in the $M_{\text{BH}}\text{-}M_*$ plane at different redshifts (panels a–h), from $z = 15$, when the Universe is younger than 300 Myr, to $z = 0$. The $M_{\text{BH}}\text{-}M_*$ relation (black curves) is obtained from these galaxies at each redshift by a running median and $1\text{-}\sigma$ (16th–84th) percentiles of M_{BH} at each given M_* . All galaxies hosted by the

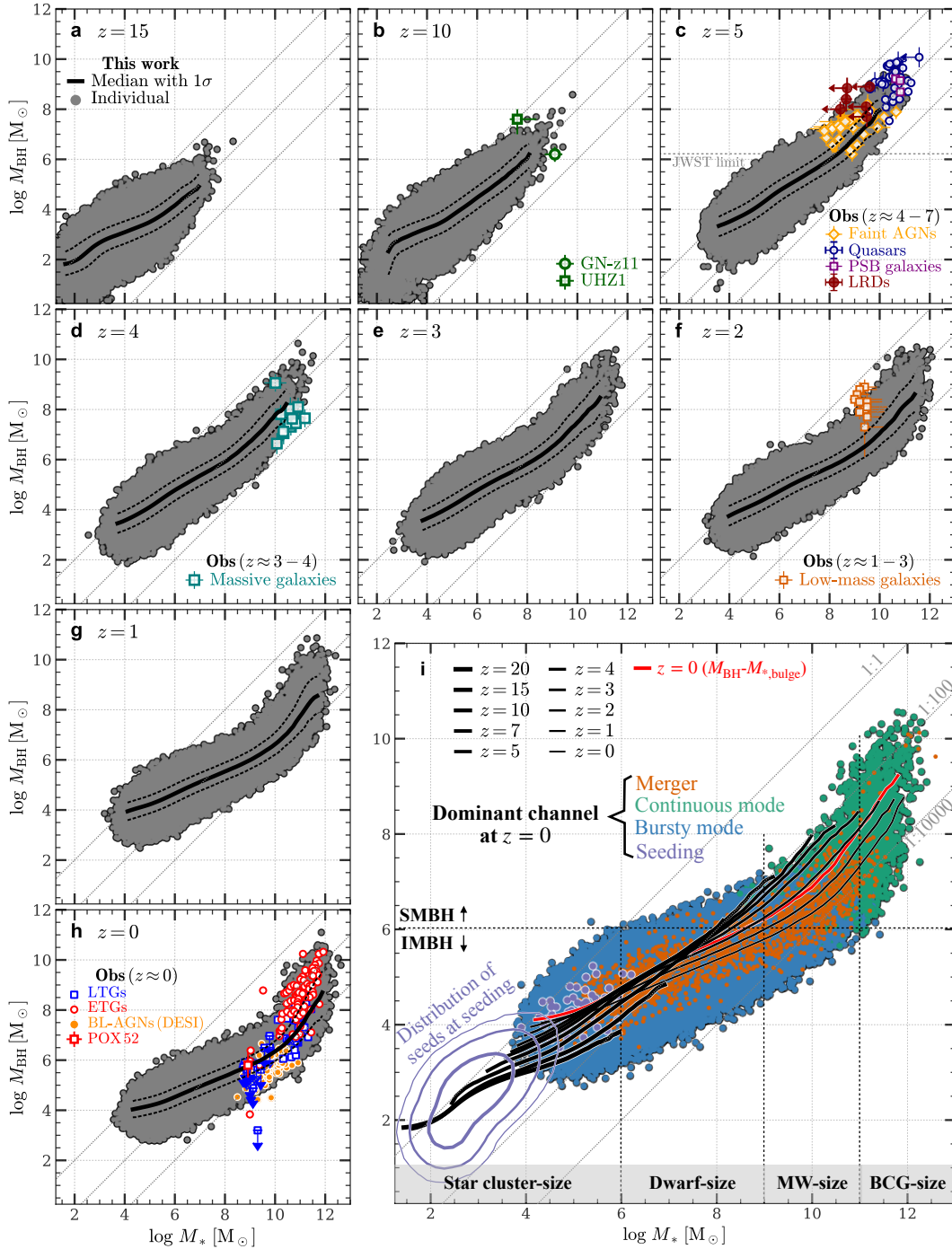


Figure 17. $M_{\text{BH}}-M_*$ relations. **a–h**, each showing the relation at a given redshift, as labeled. **Grey dots** represent individual galaxies predicted by the model, and **black curves** are the running median and $1\text{-}\sigma$ ($16^{\text{th}}\text{--}84^{\text{th}}$) percentiles at given values of M_* . **i**, a summary of the predicted median $M_{\text{BH}}-M_*$ relations at different redshifts (**black curves**). For comparison, we also show predicted median black hole mass-stellar bulge mass ($M_{\text{BH}}-M_{*,\text{bulge}}$) relation at $z=0$ (**red curve**), and the distribution of BH seeds with their M_{BH} and M_* evaluated at the seeding epochs (**purple contours**; from inner to outer, enclosing 68%, 95% and 99.7% of the seeds, respectively). Dots represent galaxies at $z=0$, with color of each indicating the channel that contributes the most to the M_{BH} (**purple**, seeding; **blue**, burst mode; **green**, continuous mode; **orange**, merger). **Colored markers** in **a–h** show observations obtained by different studies: sources at $z \approx 10\text{--}11$ obtained by Maiolino et al. (2024a, GN-z11) and Bogdán et al. (2023, UHZ1); faint AGNs at $z \approx 4\text{--}7$ from JWST observations compiled by Pacucci et al. (2023), including the samples from Harikane et al. (2023b), Übler et al. (2023) and Maiolino et al. (2024b); ‘Little red dots’ measured by Chen et al. (2025b); objects at $z \approx 4\text{--}7$ compiled by Onoue et al. (2024), including post-starburst (PSB) galaxies hosting quasars and broad-line AGNs (Carnall et al. 2023) at $z \approx 4\text{--}7$, quasars (Izumi et al. 2021; Stone et al. 2023, 2024; Yue et al. 2024a); massive galaxies hosting normal-mass BHs $z \approx 3\text{--}4$, measured by Li et al. (2025a); low-mass galaxies hosting over-massive, broad-line AGNs at $z \approx 1\text{--}3$, reported by Mezcua et al. (2024); late-type galaxies and early-type galaxies at $z \approx 0$ compiled by Greene et al. (2020); DESI broad-line AGN candidates in the local Universe obtained by Pucha et al. (2025) with confident visual inspection for the spectra; POX 52, a local early-type dwarf with both broad line and reverberation evidences of IMBH (Kawamuro et al. 2024; Sun et al. 2025). A horizontal dashed line in **c** indicates the detection limit of JWST at $z \approx 5$. In all panels, the tilted grey lines indicate constant ratios of M_{BH}/M_* . A log-normal noise with a standard deviation of 0.3 dex is added to the M_{BH} and M_* of each galaxy to mimic the observational uncertainties. This figure demonstrates that the $M_{\text{BH}}-M_*$ relation is a redshift-dependent, multi-piece relation, with each piece driven by a distinct channel of galaxy formation. See §5.4 for the details.

TNG100-1-Dark subhalos in our sample (see §2) are included in the analysis. A log-normal noise with a standard deviation of 0.3 dex is added to both M_{BH} and M_* of each galaxy to mimic observational uncertainties (e.g. Behroozi et al. 2013; Kormendy & Ho 2013). For comparison, we collect observational results in the literature at redshifts similar to those of the modeled relations, and show them in the corresponding panels. In panel i, we summarize the build-up of the $M_{\text{BH}}-M_*$ relation predicted by our model by showing (i) the distribution of seeds (purple contours), with their M_{BH} and M_* evaluated at their seeding epochs; (ii) a collection of the modeled relations at different redshifts (black curves with varying line-widths); (iii) the distribution of individual galaxies at $z = 0$, color coded according to the channel that contributes the most to the M_{BH} at $z = 0$.

At $z = 15$, the $M_{\text{BH}}-M_*$ relation shows two branches, one with $M_{\text{BH}} : M_* \approx 1 : 1$, and the other with $M_{\text{BH}} : M_* \approx 1 : 100$, connected by a smooth transition around $M_* \approx 10^4 M_\odot$. Galaxies only occupy the lower left quadrant of the $M_{\text{BH}}-M_*$ plane, all falling into the ranges of star cluster-size and dwarf-size galaxies ($< 10^9 M_\odot$, indicated in Fig. 17i), with a few extremely massive outliers reaching $M_* \approx 10^9$ and $M_{\text{BH}} \gtrsim 10^7 M_\odot$. At $z \gtrsim 10$, the two branches persist but more galaxies have evolved to the upper boundary of dwarf-size range ($M_* = 10^9 M_\odot$). The galaxy GN-z11 at $z = 10.60$ (Tacchella et al. 2023; Maiolino et al. 2024a; Ji et al. 2025) falls into the most massive tip of the galaxy distribution, with an under-massive BH touching the lower boundary of M_{BH} of the modeled galaxies with similar M_* . The galaxy UHZ1 at $z \approx 10.32$ reported by Bogdán et al. (2023) falls onto the other side of the same tip, with an over-massive BH touching the upper boundary of M_{BH} of the modeled galaxies with similar M_* . At $z = 5$ and 4, the branch with $M_{\text{BH}} : M_* \approx 1 : 1$ fades away, leaving only a tail that joins the low-mass end of the other branch. Observed galaxies at these redshifts exhibit significant diversity in the $M_{\text{BH}}-M_*$ plane: at a given M_* , the observed M_{BH} can vary by about 3 dex. The most massive galaxies in our sample have $M_* \gtrsim 10^{10} M_\odot$ and $M_{\text{BH}} \gtrsim 10^8 M_\odot$, overlapping with range covered by bright quasars (Izumi et al. 2021; Stone et al. 2023, 2024; Yue et al. 2024a) and massive post-starburst galaxies (Carnall et al. 2023; Onoue et al. 2024) found in observations. Less massive galaxies with $10^8 \lesssim M_*/M_\odot \lesssim 10^{10}$ partially overlap with the reported values of M_{BH} and M_* for faint AGNs (Pacucci et al. 2023; Harikane et al. 2023b; Übler et al. 2023; Maiolino et al. 2024b), LRDs (Chen et al. 2025b) and massive galaxies hosting normal-mass BHs (Li et al. 2025a), indicating that each sample contains a biased sub-population of galaxies. At $z \lesssim 3$, the survived branch starts to deviate from the line of $M_{\text{BH}} : M_* \approx 1 : 100$. Such a deviation is mass-dependent, with galaxies around the MW-mass ($M_* \approx 10^{10} M_\odot$) shifted most significantly to the right. The observed low-mass galaxies with $M_* \lesssim 10^{10} M_\odot$ at $z \approx 1-3$ (Mezcua et al. 2024) appear to overlap with the upper boundary of the $M_{\text{BH}}-M_*$ scatters. At $z \lesssim 1$, the $M_{\text{BH}}-M_*$ relation stabilizes as a down-bending, bow-shaped curve. The observed ETGs and LTGs compiled by Greene et al. (2020) occupy the high- M_* end and the knee of the bow with M_{BH} above and below the median relation at the corresponding M_* , respectively. The broad-line AGN candidates in low-mass galaxies obtained by Pucha et al. (2025) from Dark Energy Spectroscopic Instrument (DESI) survey, and POX 52, a local early-type dwarf with evidences for IMBH from both broad line and MIR-optical lag (Kawamuro et al. 2024; Sun et al. 2025), also occupy the knee of the relation.

The evolution of the $M_{\text{BH}}-M_*$ relation shown in Fig. 17i confirms our expectation that no single power-law relation can describe the relation at any redshift, and the broad consistency between model predictions and observations reinforces the plausibility of the physical

processes implemented. By covering the entire lifetimes of galaxies and their SMBHs, our model is capable of recovering and predicting the $M_{\text{BH}}-M_*$ relation over the entire ranges of M_{BH} and M_* and across the entire cosmic history relevant to galaxy formation. More importantly, we are now able to dissect the relation into pieces, and reveal the underlying physical mechanisms that shape each piece, as to be described below.

5.4.2 Physical processes shaping the relation

As introduced in the beginning of §5.4, the $M_{\text{BH}}-M_*$ relation at a given redshift z can be understood by dissecting the galaxy population at z into sub-populations with similar final halo mass ($M_{\text{V},z=0}$), identifying the channel that dominates the growth of each sub-population at z , and examining the $M_{\text{BH}}-M_*$ relation shaped by each channel at z (§5.3.1 and Fig. 13).

During the seeding era ($z \approx 20-30$, when most seeds are bred; see §5.3.2 and Fig. 10), BH seeds are bred with $M_{\text{BH}} \sim M_*$ due to the top-heavy IMF of Pop-III stars adopted by our model (see Eq. 34 and Fig. 4). This sets the starting loci of the $M_{\text{BH}}-M_*$ relation, and produces the highest ratio, $M_{\text{BH}}/M_* \approx 1$, that can be achieved ever in the growth history of galaxies (Fig. 17i, purple contours). All the BH seeds are bred within the regime of IMBHs, waiting for the post-seeding growth to boost them to SMBHs (i.e. move across the horizontal dashed lines in Fig. 17i).

The post-seeding growth of galaxies starts with the burst mode, which consists of nuclear bursts within the SNF nuclei formed due to global disturbances of each galaxy. The value of M_{BH}/M_* can be kept around 1/10 when the burst mode dominates the growth (Fig. 13). Owing to the super-Eddington accretion in this mode, the masses of the BHs can be boosted to $\gtrsim 10^6 M_\odot$ by only a few nuclear bursts (Fig. 14, rows 3 and 4). The bursty mode turns out to be the key condition that raises the masses of all BHs, especially those low-mass seeds, to the regime of SMBHs ($M_{\text{BH}} \gtrsim 10^6 M_\odot$), so that the subsequent growth of BHs in the continuous mode is initialized with sufficiently high M_{BH} to ensure that the galaxy is well-regulated by the AGN feedback so as to sustain a $M_{\text{BH}}-M_*$ relation similar to that observed at low z . The bursty mode thus provides a physical explanation for the origin of the ‘enhancement factor’, inferred phenomenologically in Paper-I (see their fig. 7 and §4.3.3), that boosts the BHs in low-mass halos to the regime of SMBHs. Following the bursty era, the growth of BHs via nuclear bursts is self-limited by the feedback, and the continuous-mode growth gradually takes over and keeps the value of M_{BH}/M_* around 1/100 (§5.3.1 and Fig. 13).

The $M_{\text{BH}}-M_*$ relation at $z \gtrsim 10$ can thus be described as a composition of three components, each driven by one of the above channels: the branch forming the low-mass end (with the lowest $M_{\text{V},z=0}$ among all) is populated with galaxies just seeded with BHs and has the highest $M_{\text{BH}}/M_* \approx 1$; the other branch forming the high-mass end (with the highest $M_{\text{V},z=0}$ among all) is populated with galaxies already transited to the continuous mode and has $M_{\text{BH}}/M_* \approx 1/100$; the smooth connection between the two branches is populated by galaxies in the bursty mode, with both $M_{\text{V},z=0}$ and M_{BH}/M_* in between.

Towards lower redshifts, the efficiency of AGN feedback at a given M_{V} increases due to the shallower potential (quantified by V_{max} of the halo) and stronger coupling between feedback energy and the SGC (modeled by the factor α_{AGN} ; see Appendix A for details). As a result, growth of BHs in dynamically hot SGCs at lower redshift exhibit stronger self-regulation, leading to a $M_{\text{BH}}-M_*$ relation with a lower amplitude down to $M_{\text{BH}}/M_* = 1/1000$, as shown by the red curve in Fig. 17i at $z = 0$, for which the masses of stellar bulges are

used to bypass the effect of dynamical transition (see below). The relation shows a logarithmic slope greater than 1 in the high-mass end, as expected from arguments of binding energy adopted in the modeling of AGN feedback (see §4.3 of [Paper-I](#) for details). The net effect is that the $M_{\text{BH}}-M_*$ relation starts to bend downwards at $z \lesssim 5$.

At $z \approx 2$, halos hosting the progenitors of MW-size galaxies (with $M_{\text{V},z=0} \approx 10^{12} M_{\odot}$ or $M_{*,z=0} \approx 10^{10.5} M_{\odot}$) start to transit from the fast phase to the slow phase (see Fig. 1b). Galaxies there start to efficiently grow their dynamically cold (disk) components in the continuous mode, but the accretion by BHs is terminated almost completely (§5.3.1 and Fig. 13). These MW progenitors thus start to migrate horizontally in the $M_{\text{BH}}-M_*$ plane, and eventually reach the knee of the bow-shaped $M_{\text{BH}}-M_*$ relation at $z = 0$. Halos hosting less massive galaxies (e.g. dwarfs with $M_{*,z=0} \leq 10^9 M_{\odot}$) transit to the slow phase earlier (Fig. 1b), but the post-transition star formation is less efficient than that in the MW-size progenitors due to their shallower gravitational potentials to resist feedback, so that the amount of horizontal migration in the $M_{\text{BH}}-M_*$ plane is smaller. Halos hosting more massive galaxies (e.g. the BCG-size galaxies with $M_{*,z=0} \geq 10^{11} M_{\odot}$) transit to the slow phase later. Both the star-formation efficiency and the duration left for the star formation in the dynamically cold phase in these halos are smaller than those in MW-size galaxies, again making the amount of horizontal migration in the $M_{\text{BH}}-M_*$ plane smaller. The mass-dependence of the halo transition time and the post-transition star-formation efficiency thus gradually ‘draw the bow’ at $z \lesssim 3$, and eventually build up the bow-shaped, down-bending $M_{\text{BH}}-M_*$ relation at $z = 0$, as seen by a comparison between the red and black curves in Fig. 17i at $z = 0$. Note that dry mergers of galaxies are not important in most parts of the $M_{\text{BH}}-M_*$ relation, since the mass added by each merger only takes a temporary dominance in a short period after the merger and is soon surpassed by the growth via other channels (see the brown segments in Fig. 13). The exception is in the high-mass end ($M_* \gtrsim 10^{11} M_{\odot}$) of the relation at low z ($z \lesssim 2$), where repeated mergers in each galaxy tend to drive galaxies towards a relation of $M_{\text{BH}} \propto M_*$ (produce the scatter in the high-mass tip in Fig. 17i; see also Fig. 13d).

As a summary, our model predicts a redshift-dependent $M_{\text{BH}}-M_*$ relation that is smoothly joined by multiple pieces at each redshift. The growth of galaxies within each piece, depending on the halo mass and redshift, is driven by a distinct channel, thus building up a relation with distinct amplitude and slope. The snapshot views provided by observations, although still incomplete and subject to various uncertainties, appear to support the model predictions. The $M_{\text{BH}}-M_*$ relations of the ETGs and LTGs at $z \approx 0$ were found to be distinct at $M_* \gtrsim 10^{10} M_{\odot}$, with the former showing higher amplitude and shallower slope than the latter ([Graham & Sahu 2023](#); see also §5.3.1 and Fig. 13), consistent with the multi-piece prediction of our model. At $M_* \sim 10^9 M_{\odot}$, the faint AGNs (including LRDs) at $z \approx 4-7$ have M_{BH} higher than those expected from the relation at $z \approx 0$, supporting that the $M_{\text{BH}}-M_*$ relation evolves with redshift. The relation in the regime of dwarf-size galaxies ($M_{*,z=0} \leq 10^9 M_{\odot}$) turns out to be a missed part, and is expected to be the key to test the predictions of the model and to constrain the physical processes proposed to drive the evolution. The upper bounds for the masses of IMBHs reported by [Greene et al. \(2020\)](#), and the masses of BL-AGN candidates reported by [Pucha et al. \(2025\)](#), appear to align with the predicted $M_{\text{BH}}-M_*$ distribution of galaxies at $M_{*,z=0} \leq 10^{10} M_{\odot}$ (Fig. 17h), providing some evidence to support our model. The most intriguing cases are halos that host galaxies with $M_{*,z=0} \leq 10^6 M_{\odot}$ (comparable to those of star clusters). This population can have M_{BH}/M_* comparable to or even higher than those of BCG-size galaxies (Fig. 17i). Their distribution in the $M_{\text{BH}}-M_*$ plane at

$z = 0$ aligns with the massive end of the distribution of seeds at their seeding epochs, with a few cases having M_{BH} dominated by the seeds (purple dots). The post-seeding growth of these star cluster-size galaxies is the least important among all galaxies, and observational searching for BHs in this population of galaxies may provide crucial data to test physical processes directly related to BH seeds (see §3).

6 SUMMARY AND DISCUSSION

In this paper, we have developed a model for the formation and evolution of (SM)BHs with their host galaxies and halos within the two-phase framework of galaxy formation established in [Paper-I](#). We focus particularly on the processes in galaxies during the seeding era ($z \gtrsim 20$), in which BH seeds with a broad mass spectrum and distinct ‘flavors’ are bred out of the first-generation (Pop-III) stars formed in mini-halos ($T_{\text{V}} \approx 10^3-10^4$ K), and the bursty era ($z \approx 5-20$), in which the growth of BHs is dominated by nuclear bursts within supernova-free (SNF) nuclei formed in galactic centers. The model covers wide ranges of spatial, temporal and mass scales ($r \approx 10^{-3}-10^6$ pc, Fig. 6; $z \approx 40-0$, Fig. 11; $M_{\text{BH}} \approx 10-10^{10} M_{\odot}$, Fig. 16), so that the entire lifetimes of SMBHs and their host galaxies can be self-consistently initialized and evolved within a cosmological context. We have detailed the physical processes that seed (§3) and grow (§4) SMBHs with their host galaxies and halos, derived equations describing each process, and presented our analytical approximations so that one can obtain some order-of-magnitude estimates for the processes without implementing the full model. Our main findings and conclusions are summarized as follows.

- (i) The assembly of dark matter halos provides a condition necessary for the condensation of gas that gives birth to the self-gravitating gas clouds (SGCs), and determines the locations and redshifts at which galaxies start their lives. Compact Pop-I, II and III star clusters are expected to form in different, but overlapping, regimes in the $M_{\text{V}}-z$ plane, relying on the baryonic environment to fully disentangle the overlapping. Specifically, Pop-III star clusters can form in mini-halos within a narrow band in the $M_{\text{V}}-z$ plane ($T_{\text{V}} \approx 10^3-10^4$ K), provided that self-enrichment and the enrichment from the intergalactic medium (IGM) are not present (§2; Fig. 1).
- (ii) Gas within a pristine mini-halo can collapse to form a Pop-III star cluster once the halo reaches the H_2 -cooling limit ($T_{\text{V}} \approx 10^3$ K). Two important heating sources, the Lyman-Werner (LW) radiation produced by star-forming galaxies and the dynamical heating from the fast assembly of the host halo, can delay the gas collapse up to the H-cooling limit ($T_{\text{V}} \approx 10^4$ K) and raise the mass of the star cluster generated (§3.1; Fig. 3).
- (iii) The radiative feedback produced by the dominant star in a Pop-III star cluster regulates the accretion of the star itself. Thus, a feedback-regulated model is adopted to derive the mass of the dominant star. This, together with the assumed top-heavy IMF, gives the mass spectrum of Pop-III stars. Depending on the mass and size of the host SGC, a dominant star with Eddington-level accretion can grow to a mass higher than that only with sub-Eddington accretion. The evolution of the dominant star depends on its mass, and can produce either a core-collapse supernova (CCSN), a pair-instability supernova (PISN), or a direct-collapse black hole (DCBH), from which a BH seed is bred and the mass of the seed is determined. On the other hand, a Pop-II seed with a low mass is bred if the IGM has already been enriched (§3.2; Fig. 4).
- (iv) The mass and ‘flavor’ (Pop-II or Pop-III; Eddington or sub-Eddington; CCSN, PISN or DCBH; with or without strong LW radiation or dynamical heating) of a BH seed is determined by the

cosmological context (i.e. environment) composed of the mass and assembly history of the host halo, and the radiation and matter feedback from other galaxies (Fig. 2).

- (v) To resolve the wide range of scales required to model the post-seeding growth of BHs, we use a ladder-like profile to characterize the density around each BH. The ladder consists of five steps, each describing one of the five levels in the cosmic hierarchy: the Universe, halos, galaxies (SGCs), galactic nuclei, and the accretion disks of BHs. Equations governing each level and relating one level to the next are derived, so that one can start from the outer boundary and ‘climb up’ the ladder to obtain the gas structure deep into the vicinity of a BH (§4.1; Fig. 6).
- (vi) A global disturbance on an SGC is expected to drive it out of equilibrium, overpower the consumption of gas in the galactic center, and cause the formation of an SNF nucleus with a density of $n_g \gtrsim n_{\text{snf}} = 10^{3.5} \text{ cm}^{-3}$ (i.e. $t_{\text{ff}} \lesssim 1 \text{ Myr}$) and a mass of $M_{\text{nuc}} \approx 10^7 - 10^8 M_{\odot}$ (for the progenitors of MW-size galaxies at $z \approx 10$). Our model assumes that a temporary excursion of halo assembly to a high specific growth rate (γ_v) causes such a disturbance. Once the SNF nucleus has formed, the formation of a compact stellar system and significant growth of the central BH seem inevitable, proceeding rapidly within a timescale of $\lesssim 1 \text{ Myr}$ and manifesting itself as a nuclear burst (§4.2.1).
- (vii) The strong inflow during the formation of an SNF nucleus, potentially aided by the ‘positive stellar feedback’, is expected to drive supersonic turbulence that supports the structure of the nucleus and shapes a density profile different from that in the outer volume of the host galaxy. We parameterize the density profile of an SGC using a double power-law function that smoothly connects the parts outside and inside the radius of the nucleus (R_{nuc}). The structure properties of an SNF nucleus turn out to be modulated by the accretion of the host halo, allowing the cosmological assembly on halo scales ($r \gtrsim 100 \text{ kpc}$ for $M_V \gtrsim 10^{11} M_{\odot}$ at $z = 0$) to propagate all the way down to regions close to the central BH ($r \lesssim$ a few pc for $M_{\text{BH}} \lesssim 10^6 M_{\odot}$; §4.2.2).
- (viii) The turbulent motion of gas plays a key role in the growth of BHs, and the effects of turbulence can be summarized as a ‘deferred mixing’. In the outer part of an SGC, dense and ballistically moving sub-clouds are produced in the turbulent gas, making the dissipation via collision between sub-clouds and drag force on them negligible. A broad angular momentum (j) distribution is thus preserved, and sub-clouds with low j can be captured continuously by the central BH to feed a ‘continuous mode’ of growth (see Paper-I). Associated with the formation of an SNF nucleus, sub-clouds approaching the inner SGC can be affected strongly by the collision and drag force. This leads to efficient dissipation and mixing of the gas, which then enhances the BH growth in a nuclear burst and promotes the formation of a nuclear star cluster (NSC). A turbulence-modified Bondi accretion model is used to describe the accretion of gas across the Bondi radius in such a ‘bursty mode’, which predicts super-Eddington growth rates in a wide range of parameter space (§4.3).
- (ix) During a nuclear burst, a turbulent and magnetized ($v_t \sim v_A \gg c_s$) accretion disk is expected to form around the BH. The disk is modeled according to the analytic model of Hopkins et al. (2024c), with a boundary condition set to match our structure equations for the SNF nucleus. Such a disk is featured by a thickened geometry, a large inflow rate, and a large Toomre- Q parameter, all originated from the strong turbulence and magnetic field. Consequently, the disk is expected to be stable against fragmentation, be free of star formation, and sustain the super-Eddington Bondi flow across the zone of influence of the BH (§4.4).
- (x) During each nuclear burst, a refined treatment is proposed to trace

the evolution of the the nucleus, including star formation, BH accretion, and feedback from them, so that individual bursts in the formation history of any galaxy can be resolved with a time step of the order of one year. The basic equation is built on the conservation of mass, with a number of sink terms describing the sources that deplete the gas in the nucleus. Our analyses based on depletion timescales show that AGN feedback dominates gas depletion in most cases, while stellar feedback is not important due to the ‘supernova-free’ nature of the nuclei. BH accretion competes with star formation during the nuclear burst, outpacing (or being outpaced by) star formation when the BH mass is significantly higher (or lower) than the gas mass of the nucleus (§4.5; Fig. 7).

We have implemented the above processes into subhalo merger trees constructed from a N-body simulation, presented our model predictions, and demonstrated the potential applications to be explored in the future (§5). The pipeline of the model is schematically shown in Fig. 8, with a list of the steps outlined in §5.1 and detailed in Appendix A. Our main results are summarized as follows.

- (i) BH seeds have a broad mass spectrum covering $\sim 10 - 10^5 M_{\odot}$. The mass function of Pop-III seeds shows significant bimodality, with a valley at $M_{\text{BH}} \approx 10^2 M_{\odot}$ created by PISNe, a low-mass peak at about $10^{1.7} M_{\odot}$ populated mostly with seeds bred from CCSNe and PISNe, and a high-mass peak at about $10^{2.5} M_{\odot}$ populated mostly with DCBHs. LW radiation and dynamical heating have a divergent effect on the masses of seeds, producing both high-mass and low-mass bumps with DCBH and Pop-II flavors, respectively, in the mass function of seeds (§5.2.1; Fig. 9).
- (ii) The cosmic seeding history shows that most BH seeds are bred at $z = 20 - 30$ and most Pop-II seeds are bred at $z = 10 - 20$. The seeding rate becomes lower at $z \lesssim 10$ and negligibly small at $z \lesssim 5$, making observational search for ongoing seeding activity at $z \lesssim 5$ challenging. The two environmental effects, LW radiation and dynamical heating, are found to act mainly on mini-halos at different epochs, $z \lesssim 20$ and $z \gtrsim 10$, respectively. Most seeds bred at $z \lesssim 10$ are delayed by LW radiation, highlighting the importance of radiation feedback in producing late-formed Pop-III seeds that are targets of some recent JWST programs and future surveys (§5.2.2; Fig. 10).
- (iii) A ‘seeding atlas’ is constructed to show the distributions of seeds with different properties (M_{BH} , Z_{IGM} , J_{LW} and γ_v) in the $M_V - z$ plane, giving clues about when and how seeds in individual halos are bred. The atlas can be divided into three distinct regimes, each dominated by one environmental process: a fusiform area shaped by dynamical heating, a tip-like structure shaped by LW radiation, and a two-component stripe shaped by IGM enrichment. Seeds with different masses and affected by different environmental processes have different and persistent clustering patterns, which may be used to retrieve information about the seeding histories for galaxies at low z (§5.2.3; Fig. 11).
- (iv) The post-seeding BH growth and star formation are driven by distinct channels, including the bursty mode, the continuous mode, and the merger. These channels sequentially emerge and become dominant successively in the growth history of any given galaxy, provided that the galaxy can reach the mass scale relevant to a channel at some epoch and its dynamically hot state is sustained to the epoch. The bursty mode dominates the growth of BHs soon after their seeding ($z \approx 20 - 30$ in most cases). The transition from the bursty mode to the continuous mode takes place at $z \approx 5 - 15$ when M_{BH} reaches about $10^6 M_{\odot}$, with more massive halos transiting earlier. The evolution of galaxies during the continuous stage can be described by the two-phase framework established in Paper-I, with transition from dynamically hot (bulge) phase to dynamically cold (disk) phase de-

terminated by the transition of the halo assembly from the fast to slow phase. Mergers become important only in BCG-size galaxies at $z \lesssim 2$ (§5.3.1; Figs. 13 and Fig. 14).

- (v) The evolution of a galaxy during its bursty stage has two distinct features, one is the back-and-forth switching between star-forming and quiescent states, and the other is the formation of a compact, globular-like nuclear star cluster. Our model therefore predicts that the formation of the first generation of quiescent galaxies can be as early as $z \approx 20$ – 30 , soon after the seeds are bred, and that quiescent galaxies and galaxies with extremely compact morphology are ubiquitous at $z \gtrsim 10$ (§5.3.1).
- (vi) The cosmic histories of BH accretion and star formation are compositions of those of individual galaxies. As a result, the cosmic histories are also driven, sequentially, by the above channels, and can be divided into three eras: the seeding era ($z \approx 20$ – 30), the bursty era ($z \approx 5$ – 20), and the continuous era ($z \lesssim 5$). The evolution of cosmic star formation rate density and BH accretion rate density predicted by our model are broadly consistent with JWST observations (§5.3.2; Fig. 15).
- (vii) A set of idealized and controlled experiments is performed with two variants of our model, ‘MinSeed’ and ‘MaxSeed’, both adapted from the Default model with extreme assumptions that minimize and maximize the masses of seeds, respectively. The self-regulated nature of BH accretion tends to erase the memory of the seeding method, making the differences in M_* and M_{BH} at low z between the two variants indistinguishable. Halos with $M_{\text{v},z=0} \lesssim 10^{11} M_{\odot}$ are the only population whose difference in M_{BH} caused by the seeding method can persist to $z \lesssim 8$ (§5.3.3; Fig. 16).
- (viii) The ensemble of BHs and galaxies, and their growth histories predicted by our model allow the archaeology and futurology to link observed BHs and galaxies across the cosmic history. We demonstrate the idea by comparing the predicted evolution of M_{BH} and M_* to the BHs and galaxies observed by JWST. The bright galaxies (massive BHs) observed at $z \gtrsim 10$ and the bright quasars at $z \gtrsim 6$ seem to align with the predicted evolution of galaxies hosted by massive halos with $M_{\text{v},z=0} \approx 10^{15} M_{\odot}$, and are expected to evolve to be the massive BCGs in Coma-like clusters in the local Universe. The faint AGNs (including little red dots) observed at $z \gtrsim 4$ appear to be hosted by halos with $M_{\text{v},z=0} \approx 10^{13}$ – $10^{14} M_{\odot}$. The origins and fates of a few low-mass galaxies with extraordinary morphology and enrichment pattern at $z \gtrsim 4$ are also inferred using our model (§5.3.4; Fig. 16).
- (ix) Our model predicts a M_{BH} – M_* relation that covers the full lifetimes of galaxies (from the seeding era, $z \gtrsim 20$ to present day) and the full spectrum of mass (from star cluster-size galaxies with $M_* < 10^6 M_{\odot}$ to BCG-size galaxies with $M_* \gtrsim 10^{11} M_{\odot}$). The predicted relation is a redshift-dependent, multi-piece relation, with each piece shaped by one of the growth channels. BH seeds populate the low-mass end, $M_* \lesssim 10^5 M_{\odot}$, of the relation, with $M_{\text{BH}}/M_* \approx 1$. The bursty mode provides the key condition to raise the mass of all BH seeds to the regime of SMBHs, and can sustain a relation of $M_{\text{BH}}/M_* \approx 1/10$ at $M_{\text{BH}} \lesssim 10^6 M_{\odot}$. The continuous mode takes over at higher mass, and can sustain a relation of $M_{\text{BH}}/M_* \approx 1/100$ during the dynamically hot phase of galaxies. The transition of galaxies to the dynamically cold phase prevent BHs from growing further and drive a horizontal migration in the M_{BH} – M_* plane. Such a migration is mass-dependent, with a maximum at the MW mass that reaches $M_{\text{BH}}/M_* \approx 1/10^4$, thus drawing the relation to a down-bending, bow-shaped one at $z \lesssim 3$. A comparison with observations shows that the prediction of the model recovers the observed M_{BH} – M_* relation from $z \approx 10$ to $z \approx 0$ (§5.4; Fig. 17).

The model developed in this paper provides a framework to examine the co-evolution pattern of BHs and their host galaxies and halos across cosmic time, to distinguish the physical processes driving the evolution in different regimes (redshift, mass and assembly and halo, and environment), and to interpret the origin and predict the fate of observed BHs and galaxies. Clearly, the model is not definitive; it leaves many interesting questions to be addressed in the future using both theoretical and observational approaches. Here we discuss a few of them.

The environment provided by dark matter halos sets the fundamental context of galaxy formation in our model, and in most of other empirical and semi-analytical models. However, the dynamical nature of halos and the effects of their dynamics, have not been fully explored. The upper limit of the amount of baryon available for galaxy formation in a halo is assumed to be f_b (the cosmic baryon fraction) times the mass of the halo, and is used to determine the properties of the SGC hosted by the halo (Eqs. 9–13). This assumption is certainly not valid in every cases, especially for halos that interact strongly with other halos or with dense environments. The bullet cluster (e.g. Clowe et al. 2004; Markevitch et al. 2004) and dark matter-deficient galaxies (Guo et al. 2020; Lee et al. 2021; Ogiya et al. 2022; Moreno et al. 2022; Contreras-Santos et al. 2024) observed in the local Universe provide two sets of examples for high- and low-mass halos, respectively, in which dark matter and baryon gas are detached due to interactions between halos and their environments so that the baryon fraction in the structure is actually higher than f_b . The situation is expected to be more common at high z when mergers between halos are frequent (Chen et al. 2023b) and the stream velocity between dark matter and baryons is significant compared to the small virial velocities of mini-halos (Stacy et al. 2011; Maio et al. 2011; Greif et al. 2011). It is unclear what fraction of halos with different masses and at different redshifts have elevated baryon contents and how such elevations change the seeding and early growth of BHs and galaxies. Detailed analyses using hydro simulations may provide some clues.

The collapse of gas within mini-halos is determined by environmental conditions on different scales. Our model includes the radiation and matter feedback from star formation to the large-scale environment, and the dynamical heating from the fast assembly of host halos. However, the modeling is achieved with simplified treatments that ignore many details. For example, the LW emission is assumed to be isotropic around the source galaxy, and the propagation of the radiation is assumed to depend only on redshift and the distance from the source to the target. In reality, the ISM, CGM and IGM have complex structures, which may cause the escape and propagation of the LW radiation to be highly anisotropic and time-dependent (e.g. Pallottini et al. 2022). Radiation other than that in the LW band, such as that with energy > 0.76 eV, may detach the H^- and thus also affect the formation of H_2 (Latif & Khochfar 2019), as mentioned in the end of §3.1.3. Similar arguments also apply to metal ejection and diffusion. The effect of dynamical heating introduces additional complexity, as it is generated not only from halo assembly but also from the aforementioned stream velocity between dark matter and baryon gas. All these need to be revisited in detail using hydro simulations with the relevant processes implemented accurately.

Many questions regarding the formation, growth and evolution of Pop-III stars remain unsolved. The structure of Pop-III stars, especially supermassive stars (SMSs), can exhibit large fluctuations, making their evolution tracks and radiative feedback hard to model (Hosokawa et al. 2013). The accretion of gas onto Pop-III stars has been considered critical in modifying their mass spectrum (Omukai

et al. 2008; Schleicher et al. 2013). However, such accretion flow is usually anisotropic, exhibiting disk-like structure with gas/stellar clumps (Chon & Omukai 2020; Latif et al. 2022b) and magnetic fields amplified by turbulence (Sharda et al. 2020). Resolving these issues remains to be a gap in our knowledge of stellar evolution and requires simulations with dynamical ranges beyond current computational power.

A key assumption of our model to grow BHs is that the formation of an SNF nucleus can be triggered by some global disturbance on the SGC, and such a disturbance is associated with the temporary excursion of the halo assembly to high γ_v . In reality, the sources of disturbance may be diverse, such as a close flyby of another galaxy (Li et al. 2025b), galaxy-galaxy merger (Luo et al. 2014; Puskás et al. 2025), counter-rotation of cold stream (Danovich et al. 2015), violent disk instability (Agertz et al. 2009; Meng et al. 2019; Renaud et al. 2021), or even the tidal force induced by the misalignment between the galaxy and the interior of the host halo (Chisari et al. 2017). Detailed examinations of these processes require either simulations with large volumes to provide statistical samples or controlled experiments to disentangle the effects of different processes. Such simulations with sufficiently high resolution and with implementations of all relevant processes may verify or falsify that nuclei with densities higher than n_{snf} can indeed form as a result of the disturbance, as well as resolve the density profiles of the nuclei.

Super-Eddington accretion of BHs appears to be a natural outcome if SNF nuclei can indeed form around the BHs. Whether or not such a high accretion rate can be sustained depends on the structure (e.g. anisotropy and clumpiness) of the nucleus and the properties (geometry, magnetic field, radiation and jet feedback) of the accretion disk. The dry channel (i.e. BH-BH merger, and capture of stars) may also contribute to the growth of BHs, as explored by some simulations (Tremmel et al. 2015; Gaete et al. 2024; Fujii et al. 2024) and (semi-)analytical models (Partmann et al. 2024; Liu et al. 2024; Dekel et al. 2025), and produce observable signatures such as tidal disruption events (TDEs) and gravitational waves (GWs). These processes are still poorly understood and require further refinement of the model to resolve (e.g. Anglés-Alcázar et al. 2021; Latif et al. 2022b; Hopkins et al. 2024a). Observational constraints can also help to break the degeneracy in the model and to tighten the posterior space to be explored.

The formation and growth of nuclear star clusters (NSCs) are by-products of nuclear bursts. Intuitively, the competition between star formation and BH accretion naturally leads to an anti-correlation between the NSC-BH mass ratio ($M_{\text{NSC}}/M_{\text{BH}}$) and the mass of BH (M_{BH}), as suggested by some observations (e.g. Georgiev et al. 2016; Nguyen et al. 2018; Neumayer et al. 2020). However, the dry evolution of any NSC, such as that contributed by mergers with other star clusters (Tremaine et al. 1975; Capuzzo-Dolcetta & Mocchi 2008; Neumayer et al. 2011; Arca-Sedda & Capuzzo-Dolcetta 2014; Guillard et al. 2016; Tsatsi et al. 2017; Fahrion et al. 2022) and by the interaction between the NSC and the central BH (Antonini et al. 2015) can introduce significant contamination and needs to be modeled carefully.

The growth of BHs in the bursty mode provides a key condition for BH seeds to grow to SMBHs. However, as demonstrated in §5.3.3, the required boost in the BH mass degenerates with the seeding method, and current observations are unable to break the degeneracy. The key targets that can be used to isolate the effects of the bursty mode are galaxies at $z \gtrsim 10$, when the growth of BHs is dominated by nuclear bursts (see §5.3.2 and Fig. 15), and BHs in low-mass galaxies (e.g. dwarfs with $M_* < 10^9 M_\odot$ and especially star cluster-size galaxies with $M_* < 10^6 M_\odot$) where a large fraction of the total M_{BH} is

contributed by the bursty mode (see §5.3.1, Figs. 13, 14 and 17). In addition, low-mass halos, such as those with $M_v \lesssim 10^{9.5} M_\odot$ at their infall times, are around or below the atomic-cooling limit (Fig. 1) and thus expected to host galaxies with little or no star formation after the cosmic reionization. Detections of galaxies hosted by this population of halos may thus provide fossil information of galaxy formation at $z \gtrsim 6$. A calibration for the seeding method and the growth in the bursty mode using these targets may be achievable, and should be explored further.

ACKNOWLEDGEMENTS

YC is funded by the China Postdoctoral Science Foundation (grant No. 2022TQ0329). This work is also supported by the National Natural Science Foundation of China (NSFC, Nos. 12192224, 11733004 and 11890693) and CAS Project for Young Scientists in Basic Research (grant No. YSBR-062). YC thanks Tao Wang, Kai Wang, Ziwen Zhang, Yunjing Wu, Enci Wang, Long Wang, Qi Guo, Houzun Chen and Xi Kang for their valuable insights and discussions, and thanks Yi Mao and Chen Chen for their technical support. The authors would like to express their gratitude to the Tsinghua Astrophysics High-Performance Computing platform at Tsinghua University and the Supercomputer Center of the University of Science and Technology of China for providing the necessary computational and data storage resources that have significantly contributed to the research results presented in this paper.

The computations and presentations in this paper are supported by various software tools, including the HPC toolkits HIPP (Chen & Wang 2023)² and PyHIPP³, interactive computation environment IPYTHON (Perez & Granger 2007), numerical libraries NUMPY (Harris et al. 2020), ASTROPY (Robitaille et al. 2013; Astropy Collaboration et al. 2018, 2022) and SCIPY (Virtanen et al. 2020), the graphical library MATPLOTLIB (Hunter 2007), and the two-point correlation function calculator (Sinha & Garrison 2019, 2020)⁴. The implementation of the Monte Carlo trees is revised from the public code SATGEN⁵ (Jiang et al. 2021). This research has made extensive use of the arXiv and NASA’s Astrophysics Data System. Data compilations used in this paper have been made much more accurate and efficient by the software WEBPLOTDIGITIZER.

DATA AVAILABILITY

The code repository TWOPHASEGALAXYMODEL⁶ implements the model described in this series of papers. All data used in this paper, including data points displayed in figures and observational results used for comparison, will be distributed along with the repository.

REFERENCES

- Abazajian K., et al., 2003, *AJ*, 126, 2081
 Abell G. O., Corwin Jr. H. G., Olowin R. P., 1989, *ApJS*, 70, 1
 Agarwal B., Khochfar S., Johnson J. L., Neistein E., Dalla Vecchia C., Livio M., 2012, *MNRAS*, 425, 2854
² <https://github.com/ChenYangyao/hipp>
³ <https://github.com/ChenYangyao/pyhipp>
⁴ <https://github.com/manodeep/Corrfunc>
⁵ <https://github.com/shergreen/SatGen>
⁶ <https://github.com/ChenYangyao/two-phase-galaxy-model>

- Agarwal B., Cullen F., Khochfar S., Ceverino D., Klessen R. S., 2019, *MNRAS*, 488, 3268
- Agertz O., Lake G., Teyssier R., Moore B., Mayer L., Romeo A. B., 2009, *MNRAS*, 392, 294
- Aguado D. S., et al., 2023, *MNRAS*, 520, 866
- Ahn K., Shapiro P. R., Iliev I. T., Mellema G., Pen U.-L., 2009, *ApJ*, 695, 1430
- Akins H. B., et al., 2024, COSMOS-Web: The over-Abundance and Physical Nature of "Little Red Dots"—Implications for Early Galaxy and SMBH Assembly, doi:10.48550/arXiv.2406.10341
- Anglés-Alcázar D., et al., 2021, *ApJ*, 917, 53
- Antonini F., Barausse E., Silk J., 2015, *ApJ*, 812, 72
- Arca-Sedda M., Capuzzo-Dolcetta R., 2014, *MNRAS*, 444, 3738
- Arita J., Kashikawa N., Onoue M., Yoshioka T., Takeda Y., Hoshi H., Shimizu S., 2025, *MNRAS*, 536, 3677
- Astropy Collaboration et al., 2018, *AJ*, 156, 123
- Astropy Collaboration et al., 2022, *ApJ*, 935, 167
- Ayromlou M., Nelson D., Pillepich A., 2023, *MNRAS*, 524, 5391
- Baraffe I., Heger A., Woosley S. E., 2001, *ApJ*, 550, 890
- Barro G., et al., 2024, A Comprehensive Photometric Selection of 'Little Red Dots' in MIRI Fields: An IR-Bright LRD at $z=3.1386$ with Warm Dust Emission, doi:10.48550/arXiv.2412.01887
- Baumgardt H., Hilker M., 2018, *MNRAS*, 478, 1520
- Begelman M. C., Pringle J. E., 2007, *MNRAS*, 375, 1070
- Begelman M. C., Silk J., 2023, *MNRAS*, 526, L94
- Behroozi P. S., Wechsler R. H., Conroy C., 2013, *ApJ*, 770, 57
- Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, 488, 3143
- Behroozi P., et al., 2020, *MNRAS*, 499, 5702
- Benson A. J., Sugiyama N., Nusser A., Lacey C. G., 2006, *MNRAS*, 369, 1055
- Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, *MNRAS*, 436, 697
- Bezanson R., et al., 2022, The JWST UNCOVER Treasury Survey: Ultra-deep NIRSpec and NIRCам Observations before the Epoch of Reionization, doi:10.48550/arXiv.2212.04026
- Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition
- Boco L., Lapi A., Shankar F., Fu H., Gabrielli F., Sicilia A., 2023, *ApJ*, 954, 97
- Bogdán Á., et al., 2023, *Nature Astronomy*, 8, 126
- Bond J. R., Kofman L., Pogosyan D., 1996, *Nature*, 380, 603
- Bondi H., 1952, *MNRAS*, 112, 195
- Bondi H., Hoyle F., 1944, *MNRAS*, 104, 273
- Bonifacio P., Caffau E., François P., Spite M., 2025, *A&ARv*, 33, 2
- Bonnell I. A., Bate M. R., Clarke C. J., Pringle J. E., 2001, *MNRAS*, 323, 785
- Booth C. M., Schaye J., 2009, *MNRAS*, 398, 53
- Bouwens R. J., et al., 2023, *MNRAS*, 523, 1036
- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645
- Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, *MNRAS*, 465, 32
- Boylan-Kolchin M., 2023, *Nature Astronomy*
- Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150
- Bromm V., Loeb A., 2003, *ApJ*, 596, 34
- Brown G., Gnedin O. Y., 2021, *MNRAS*, 508, 5935
- Bruzual G., Charlot S., 2003, *MNRAS*, 344, 1000
- Bryan G. L., Norman M. L., 1998, *ApJ*, 495, 80
- Capuzzo-Dolcetta R., Mocchi P., 2008, *MNRAS*, 388, L69
- Carnall A. C., et al., 2023, *Nature*, 619, 716
- Carniani S., et al., 2024, *Nature*, 633, 318
- Casado J., Ascasibar Y., Gavilán M., Terlevich R., Terlevich E., Hoyos C., Díaz A. I., 2015, *MNRAS*, 451, 888
- Ceverino D., Dekel A., Bournaud F., 2010, *MNRAS*, 404, 2151
- Chen Y., Wang K., 2023, Astrophysics Source Code Library, p. ascl:2301.030
- Chen Y., Mo H. J., Li C., Wang H., Yang X., Zhou S., Zhang Y., 2019, *ApJ*, 872, 180
- Chen Y., Mo H. J., Li C., Wang H., Yang X., Zhang Y., Wang K., 2020, *ApJ*, 899, 81
- Chen Y., Mo H. J., Li C., Wang K., 2021a, *MNRAS*, 504, 4865
- Chen Y., Mo H. J., Li C., Wang K., Wang H., Yang X., Zhang Y., Katz N., 2021b, *MNRAS*, 507, 2510
- Chen Y., Mo H. J., Li C., Wang K., Wang H., Yang X., 2023a, *MNRAS*, 525, 1254
- Chen Y., Mo H. J., Wang K., 2023b, *MNRAS*, 526, 2542
- Chen Y., Mo H., Wang H., 2024, *MNRAS*, 532, 4340
- Chen Y., Mo H., Wang H., 2025a, *MNRAS*, 540, 1235
- Chen C.-H., Ho L. C., Li R., Zhuang M.-Y., 2025b, *ApJ*, 983, 60
- Chisari N. E., et al., 2017, *MNRAS*, 472, 1163
- Choi J.-H., Shlosman I., Begelman M. C., 2013, *ApJ*, 774, 149
- Chon S., Omukai K., 2020, *MNRAS*, 494, 2851
- Clowe D., Gonzalez A., Markevitch M., 2004, *ApJ*, 604, 596
- Conroy C., White M., 2013, *ApJ*, 762, 70
- Contreras-Santos A., Buitrago F., Knebe A., Rasia E., Pearce F. R., Cui W., Power C., Winstanley J., 2024, *A&A*, 690, A109
- Costa T., Sijacki D., Haehnelt M. G., 2014, *MNRAS*, 444, 2355
- Costa G., Mapelli M., Iorio G., Santoliquido F., Escobar G. J., Klessen R. S., Bressan A., 2023, *MNRAS*, 525, 2891
- Crain R. A., et al., 2015, *MNRAS*, 450, 1937
- Croton D. J., et al., 2006, *MNRAS*, 365, 11
- Curtis M., Sijacki D., 2016, *MNRAS*, 463, 63
- Curtis-Lake E., et al., 2023, *Nature Astronomy*, 7, 622
- D'Antona F., et al., 2023, *A&A*, 680, L19
- D'Eugenio F., et al., 2024, *Nature Astronomy*
- Dalton G. B., Maddox S. J., Sutherland W. J., Efstathiou G., 1997, *MNRAS*, 289, 263
- Danhaive A. L., et al., 2025, The Dawn of Disks: Unveiling the Turbulent Ionised Gas Kinematics of the Galaxy Population at $z \sim 4-6$ with JWST/NIRCam Grism Spectroscopy
- Danovich M., Dekel A., Hahn O., Ceverino D., Primack J., 2015, *MNRAS*, 449, 2087
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- De Lucia G., Blaizot J., 2007, *MNRAS*, 375, 2
- Dekel A., et al., 2009, *Nature*, 457, 451
- Dekel A., Zolotov A., Tweed D., Cacciato M., Ceverino D., Primack J. R., 2013, *MNRAS*, 435, 999
- Dekel A., Sarkar K. C., Birnboim Y., Mandelker N., Li Z., 2023, *MNRAS*, 523, 3201
- Dekel A., Stone N. C., Chowdhury D. D., Gilbaum S., Li Z., Mandelker N., van den Bosch F. C., 2025, *A&A*, 695, A97
- Dijkstra M., Haiman Z., Rees M. J., Weinberg D. H., 2004, *ApJ*, 601, 666
- Dijkstra M., Ferrara A., Mesinger A., 2014, *MNRAS*, 442, 2036
- Dolag K., Borgani S., Murante G., Springel V., 2009, *MNRAS*, 399, 497
- Dong F., Zhao D., Han J., Li Z., Jing Y., Yang X., 2022, *ApJ*, 929, 120
- Drakos N. E., Taylor J. E., Benson A. J., 2022, *MNRAS*, 516, 106
- Du M., et al., 2019, *ApJ*, 875, 58
- Dutta R., Sharma P., Sarkar K. C., Stone J. M., 2024, *ApJ*, 973, 148
- Eswaran V., Pope S. B., 1988, *Computers and Fluids*, 16, 257
- Fahrion K., et al., 2022, *A&A*, 667, A101
- Farmer R., Renzo M., de Mink S. E., Marchant P., Justham S., 2019, *ApJ*, 887, 53
- Federrath C., Klessen R. S., Iapichino L., Beattie J. R., 2021, *Nature Astronomy*, 5, 365
- Fernandez R., Bryan G. L., Haiman Z., Li M., 2014, *MNRAS*, 439, 3798
- Finkelstein S. L., et al., 2023, *ApJ*, 946, L13
- Fujii M. S., Wang L., Tanikawa A., Hirai Y., Saitoh T. R., 2024, *Science*, 384, 1488
- Fujimoto S., et al., 2024, Primordial Rotating Disk Composed of $\gtrsim 15$ Dense Star-Forming Clumps at Cosmic Dawn, doi:10.48550/arXiv.2402.18543
- Fujimoto S., et al., 2025, GLIMPSE: An Ultra-Faint $\sim 10^5 M_{\odot}$ Pop III Galaxy Candidate and First Con-

- straints on the Pop III UV Luminosity Function at $z \sim 6-7$, [doi:10.48550/arXiv.2501.11678](https://doi.org/10.48550/arXiv.2501.11678)
- Gaete B., Schleicher D. R. G., Lupi A., Reinoso B., Fellhauer M., Vergara M. C., 2024, Supermassive Black Hole Formation via Collisions in Black Hole Clusters, [doi:10.48550/arXiv.2406.13072](https://doi.org/10.48550/arXiv.2406.13072)
- Gao L., White S. D. M., 2007, *Monthly Notices of the Royal Astronomical Society: Letters*, 377, L5
- García F. A. B., Ricotti M., Sugimura K., Park J., 2023, *MNRAS*, 522, 2495
- Gaspari M., Sądowski A., 2017, *ApJ*, 837, 149
- Gaspari M., Brighenti F., Temi P., 2015, *A&A*, 579, A62
- Georgiev I. Y., Böker T., Leigh N., Lützgendorf N., Neumayer N., 2016, *MNRAS*, 457, 2122
- Glover S., 2013, in Wiklind T., Mobasher B., Bromm V., eds., *The First Galaxies: Theoretical Predictions and Observational Clues*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 103–174, [doi:10.1007/978-3-642-32362-1_3](https://doi.org/10.1007/978-3-642-32362-1_3)
- Graham A. W., Sahu N., 2023, *MNRAS*, 518, 2177
- Greene J. E., Zakamska N. L., Smith P. S., 2012, *ApJ*, 746, 86
- Greene J. E., Strader J., Ho L. C., 2020, *ARA&A*, 58, 257
- Greene J. E., et al., 2024, *ApJ*, 964, 39
- Greif T. H., White S. D. M., Klessen R. S., Springel V., 2011, *ApJ*, 736, 147
- Gronke M., Oh S. P., 2020, *MNRAS*, 492, 1970
- Guillard N., Emsellem E., Renaud F., 2016, *MNRAS*, 461, 3620
- Gunn J. E., Gott III J. R., 1972, *ApJ*, 176, 1
- Guo Q., et al., 2011, *MNRAS*, 413, 101
- Guo Q., et al., 2020, *Nature Astronomy*, 4, 246
- Habouzit M., Volonteri M., Dubois Y., 2017, *MNRAS*, 468, 3935
- Habouzit M., et al., 2021, *MNRAS*, 503, 1940
- Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, *MNRAS*, 375, 489
- Haiman Z., Abel T., Rees M. J., 2000, *ApJ*, 534, 11
- Harikane Y., et al., 2023a, *ApJS*, 265, 5
- Harikane Y., et al., 2023b, *ApJ*, 959, 39
- Harris C. R., et al., 2020, *Nature*, 585, 357
- He Y. Q., Xia X. Y., Hao C. N., Jing Y. P., Mao S., Li C., 2013, *ApJ*, 773, 37
- Henriques B., Yates R., Fu J., Guo Q., Kauffmann G., Srisawat C., Thomas P., White S., 2020, *MNRAS*, 491, 5795
- Hobbs A., Nayakshin S., Power C., King A., 2011, *MNRAS*, 413, 2633
- Hong H., et al., 2023, *ApJ*, 954, 183
- Hopkins A. M., Beacom J. F., 2006, *ApJ*, 651, 142
- Hopkins P. F., et al., 2023, *MNRAS*, 519, 3154
- Hopkins P. F., et al., 2024a, *The Open Journal of Astrophysics*, 7, 18
- Hopkins P. F., et al., 2024b, *The Open Journal of Astrophysics*, 7, 19
- Hopkins P. F., et al., 2024c, *The Open Journal of Astrophysics*, 7, 20
- Hosokawa T., Yorke H. W., Omukai K., 2010, *ApJ*, 721, 478
- Hosokawa T., Omukai K., Yorke H. W., 2012, *ApJ*, 756, 93
- Hosokawa T., Yorke H. W., Inayoshi K., Omukai K., Yoshida N., 2013, *ApJ*, 778, 178
- Hoyle F., Lyttleton R. A., 1939, *Proceedings of the Cambridge Philosophical Society*, 35, 405
- Hsiao T. Y.-Y., et al., 2025, SAPPHIRES: Extremely Metal-Poor Galaxy Candidates with $12 + \log(O/H) < 7.0$ at $z \sim 5-7$ from Deep JWST/NIRCam Grism Observations, [doi:10.48550/arXiv.2505.03873](https://doi.org/10.48550/arXiv.2505.03873)
- Hunter J. D., 2007, *Computing in Science & Engineering*, 9, 90
- Huško F., Lacey C. G., Schaye J., Schaller M., Nobels F. S. J., 2022, *MNRAS*, 516, 3750
- Inayoshi K., Ichikawa K., 2024, *ApJ*, 973, L49
- Inayoshi K., Haiman Z., Ostriker J. P., 2016, *MNRAS*, 459, 3738
- Iorio G., et al., 2023, *MNRAS*, 524, 426
- Izumi T., et al., 2021, *ApJ*, 914, 36
- Jackson R. A., Kaviraj S., Martin G., Devriendt J. E. G., Noakes-Kettel E. A., Silk J., Ogle P., Dubois Y., 2022, *MNRAS*, 511, 607
- Jeon J., Bromm V., Liu B., Finkelstein S. L., 2025, *ApJ*, 979, 127
- Jespersen C. K., Steinhardt C. L., Somerville R. S., Lovell C. C., 2025, *ApJ*, 982, 23
- Ji X., et al., 2025, *MNRAS*, 541, 2134
- Jia C., et al., 2025, *ApJ*, 986, L24
- Jiang F., van den Bosch F. C., 2014, *MNRAS*, 440, 193
- Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2021, *MNRAS*, 502, 621
- Jing Y. P., Mo H. J., Börner G., 1998, *ApJ*, 494, 1
- Kauffmann G., White S. D. M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Kawamuro T., Ricci C., Yamada S., Noda H., Li R., Temple M. J., Tortosa A., 2024, *ApJ*, 960, 15
- Keres D., Katz N., Weinberg D. H., Dave R., 2005, *MNRAS*, 363, 2
- Keselman J. A., Nusser A., 2012, *MNRAS*, 424, 1232
- Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, *MNRAS*, 457, 4340
- Kokorev V., et al., 2024, A Census of Photometrically Selected Little Red Dots at $4 < z < 9$ in JWST Blank Fields ([arXiv:2401.09981](https://arxiv.org/abs/2401.09981)), [doi:10.48550/arXiv.2401.09981](https://doi.org/10.48550/arXiv.2401.09981)
- Kormendy J., Ho L. C., 2013, *ARA&A*, 51, 511
- Koudmani S., Somerville R. S., Sijacki D., Bourne M. A., Jiang Y.-F., Profit K., 2024, *MNRAS*, 532, 60
- Kroupa P., Weidner C., Pflamm-Altenburg J., Thies I., Dabringhausen J., Marks M., Maschberger T., 2013, in , Vol. 5, *Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations*. p. 115, [doi:10.1007/978-94-007-5612-0_4](https://doi.org/10.1007/978-94-007-5612-0_4)
- Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, *ARA&A*, 57, 227
- Kulier A., et al., 2023, *ApJ*, 954, 177
- Labbe I., et al., 2025, *ApJ*, 978, 92
- Lacey C. G., et al., 2016, *MNRAS*, 462, 3854
- Latif M. A., Khochfar S., 2019, *MNRAS*, 490, 2706
- Latif M. A., Schleicher D. R. G., Schmidt W., Niemeyer J., 2013, *MNRAS*, 433, 1607
- Latif M. A., Khochfar S., Schleicher D., Whalen D. J., 2021, *MNRAS*, 508, 1756
- Latif M. A., Whalen D. J., Khochfar S., Herrington N. P., Woods T. E., 2022a, *Nature*, 607, 48
- Latif M. A., Whalen D., Khochfar S., 2022b, *ApJ*, 925, 28
- Lee J., Shin E.-j., Kim J.-h., 2021, *ApJ*, 917, L15
- Li H., Chen Y., Wang H., Mo H., 2024, Physical Processes Behind the Co-Evolution of Halos, Galaxies and Supermassive Black Holes in the IllustrisTNG Simulation, [doi:10.48550/arXiv.2409.06208](https://doi.org/10.48550/arXiv.2409.06208)
- Li J., Shen Y., Zhuang M.-Y., 2025a, A Prevalent Population of Normal-Mass Central Black Holes in High-Redshift Massive Galaxies, [doi:10.48550/arXiv.2502.05048](https://doi.org/10.48550/arXiv.2502.05048)
- Li F., Rahman M., Murray N., Kereš D., Wetzel A., Faucher-Giguère C.-A., Hopkins P. F., Moreno J., 2025b, *ApJ*, 979, 7
- Li Z., Inayoshi K., Chen K., Ichikawa K., Ho L. C., 2025c, *ApJ*, 980, 36
- Lin X., et al., 2023, *ApJ*, 944, L59
- Lin X., et al., 2025, *arXiv e-prints*, p. arXiv:2505.02896
- Liu H., Luo B., Brandt W. N., Brotherton M. S., Gallagher S. C., Ni Q., Shemmer O., Timlin III J. D., 2021, *ApJ*, 910, 103
- Liu B., et al., 2024, *MNRAS*, 534, 1634
- Lu Y., Mo H. J., Katz N., Weinberg M. D., 2006, *MNRAS*, 368, 1931
- Lu Z., Mo H. J., Lu Y., Katz N., Weinberg M. D., van den Bosch F. C., Yang X., 2014, *MNRAS*, 439, 1294
- Lumsden S. L., Nichol R. C., Collins C. A., Guzzo L., 1992, *MNRAS*, 258, 1
- Luo W., Yang X., Zhang Y., 2014, *ApJ*, 789, L16
- Ma X., et al., 2020, *MNRAS*, 493, 4315
- Ma W., Cui W., Davé R., Anglés-Alcázar D., Guo H., 2025a, Investigating the Residuals in the $M_{\bullet} - M_{\star}$ Relation Using the SIMBA Cosmological Simulation, [doi:10.48550/arXiv.2508.04413](https://doi.org/10.48550/arXiv.2508.04413)
- Ma Q.-B., Chen X.-R., Li M., Guo Q., Ciardi B., Acharya A., Wang X., 2025b, *ApJ*, 986, 5
- Machacek M. E., Bryan G. L., Abel T., 2001, *ApJ*, 548, 509
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Maio U., Dolag K., Ciardi B., Tornatore L., 2007, *MNRAS*, 379, 963
- Maio U., Koopmans L. V. E., Ciardi B., 2011, *MNRAS*, 412, L40
- Maiolino R., et al., 2024a, *Nature*, 627, 59
- Maiolino R., et al., 2024b, *A&A*, 691, A145
- Mapelli M., Spera M., Montanari E., Limongi M., Chieffi A., Giacobbo N., Bressan A., Bouffanais Y., 2020, *ApJ*, 888, 76
- Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, *MNRAS*, 351, 169

- Markevitch M., et al., 2004, *ApJ*, 606, 819
- Martin-Alvarez S., Devriendt J., Slyz A., Teyssier R., 2018, *MNRAS*, 479, 3343
- Matthee J., Mackenzie R., Simcoe R. A., Kashino D., Lilly S. J., Bordoloi R., Eilers A.-C., 2023, *ApJ*, 950, 67
- Matthee J., et al., 2024, *ApJ*, 963, 129
- Mayer L., van Dokkelaar F., Messa M., Capelo P. R., Adamo A., 2024, In-Situ Formation of Star Clusters at $z > 7$ via Galactic Disk Fragmentation; Shedding Light on Ultra-Compact Clusters and Overmassive Black Holes Seen by JWST, doi:10.48550/arXiv.2411.00670
- McGibbon R. J., Khochfar S., 2022, *MNRAS*, 513, 5423
- Meng X., Gnedin O. Y., 2020, *MNRAS*, 494, 1263
- Meng X., Gnedin O. Y., Li H., 2019, *MNRAS*, 486, 1574
- Meng J., Li C., Mo H. J., Chen Y., Jiang Z., Xie L., 2023, *ApJ*, 944, 75
- Merloni A., Heinz S., 2008, *MNRAS*, 388, 1011
- Mezcua M., Pacucci F., Suh H., Siudek M., Natarajan P., 2024, *ApJ*, 966, L30
- Mo H., van den Bosch F., White S., 2010, *Galaxy Formation and Evolution*. Cambridge University Press, Cambridge
- Mo H., Chen Y., Wang H., 2024, *MNRAS*, 532, 3808
- More S., Diemer B., Kravtsov A. V., 2015, *ApJ*, 810, 36
- Moreno J., et al., 2022, *Nature Astronomy*, 6, 496
- Morishita T., Liu Z., Stiavelli M., Treu T., Bergamini P., Zhang Y., 2025, Pristine Massive Star Formation Caught at the Break of Cosmic Dawn, doi:10.48550/arXiv.2507.10521
- Mowla L., et al., 2024, *Nature*, 636, 332
- Mucesh S., Hartley W. G., Gilligan-Lee C. M., Lahav O., 2024, Nature versus Nurture in Galaxy Formation: The Effect of Environment on Star Formation with Causal Machine Learning, doi:10.48550/arXiv.2412.02439
- Naidu R. P., et al., 2022, *ApJ*, 940, L14
- Naidu R. P., et al., 2024, All the Little Things in Abell 2744: $\$ > \1000 Gravitationally Lensed Dwarf Galaxies at $z=0-9$ from JWST NIRC2 Grism Spectroscopy, doi:10.48550/arXiv.2410.01874
- Naidu R. P., et al., 2025, A Cosmic Miracle: A Remarkably Luminous Galaxy at $z_{\text{spec}}=14.44$ Confirmed with JWST, doi:10.48550/arXiv.2505.11263
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Nelson D., et al., 2019, *Computational Astrophysics and Cosmology*, 6, 2
- Neumayer N., Walcher C. J., Andersen D., Sánchez S. F., Böker T., Rix H.-W., 2011, *MNRAS*, 413, 1875
- Neumayer N., Seth A., Böker T., 2020, *A&ARv*, 28, 4
- Nguyen D. D., et al., 2018, *ApJ*, 858, 118
- Niemi S.-M., 2011, Properties of Galaxies and Groups: Nature versus Nurture, doi:10.48550/arXiv.1109.3426
- Nusser A., 2024, *ApJ*, 974, 27
- O'Shea B. W., Norman M. L., 2008, *ApJ*, 673, 14
- Ogiya G., van den Bosch F. C., Burkert A., 2022, *MNRAS*, 510, 2724
- Okabe N., Futamase T., Kajisawa M., Kuroshima R., 2014, *ApJ*, 784, 90
- Omukai K., Schneider R., Haiman Z., 2008, *ApJ*, 686, 801
- Onoue M., et al., 2024, A Post-Starburst Pathway for the Formation of Massive Galaxies and Black Holes at $Z > 6$, doi:10.48550/arXiv.2409.07113
- Pacucci F., Loeb A., 2024, *ApJ*, 964, 154
- Pacucci F., Narayan R., 2024, *ApJ*, 976, 96
- Pacucci F., Nguyen B., Carniani S., Maiolino R., Fan X., 2023, *ApJ*, 957, L3
- Pallottini A., et al., 2022, *MNRAS*, 513, 5621
- Pariev V. I., Blackman E. G., Boldyrev S. A., 2003, *A&A*, 407, 403
- Parkinson H., Cole S., Helly J., 2008, *MNRAS*, 383, 557
- Parlanti E., Carniani S., Pallottini A., Cignoni M., Cresci G., Kohandel M., Mannucci F., Marconi A., 2023, *A&A*, 673, A153
- Partmann C., Naab T., Rantala A., Genina A., Mannerkoski M., Johansson P. H., 2024, *MNRAS*, 532, 4681
- Paulino-Afonso A., et al., 2019, *A&A*, 630, A57
- Peacock J. A., 2003, Large-Scale Surveys and Cosmic Structure, doi:10.48550/arXiv.astro-ph/0309240
- Peacock J. A., Smith R. E., 2000, *MNRAS*, 318, 1144
- Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton University Press
- Peng Y.-j., et al., 2010, *ApJ*, 721, 193
- Peng Y., Maiolino R., Cochrane R., 2015, *Nature*, 521, 192
- Perez F., Granger B. E., 2007, *Computing in Science and Engineering*, 9, 21
- Perger K., Fogasy J., Frey S., Gabányi K. É., 2025, *A&A*, 693, L2
- Perna M., Lanzuisi G., Brusa M., Mignoli M., Cresci G., 2017, *A&A*, 603, A99
- Perna M., Cresci G., Brusa M., Lanzuisi G., Concas A., Mainieri V., Mannucci F., Marconi A., 2019, *A&A*, 623, A171
- Pillepich A., et al., 2018, *MNRAS*, 475, 648
- Pimbblet K. A., 2011, *MNRAS*, 411, 2637
- Planck Collaboration et al., 2016, *A&A*, 594, A13
- Power C., Nayakshin S., King A., 2011, *MNRAS*, 412, 269
- Proctor K. L., Ludlow A. D., Lagos C. d. P., Robotham A. S. G., 2025, *MNRAS*
- Pucha R., et al., 2025, *ApJ*, 982, 10
- Puech M., Hammer F., Hopkins P. F., Athanassoula E., Flores H., Rodrigues M., Wang J. L., Yang Y. B., 2012, *ApJ*, 753, 128
- Puskás D., et al., 2025, *MNRAS*, 540, 2146
- Read J. I., Wilkinson M. I., Evans N. W., Gilmore G., Kleyna J. T., 2006, *MNRAS*, 366, 429
- Renaud F., Romeo A. B., Agertz O., 2021, *MNRAS*, 508, 352
- Renaud F., Agertz O., Romeo A. B., 2024, *A&A*, 687, A91
- Robitaille T. P., et al., 2013, *A&A*, 558, A33
- Rodriguez-Gomez V., et al., 2015, *MNRAS*, 449, 49
- Rosas-Guevara Y. M., et al., 2015, *MNRAS*, 454, 1038
- Rowland L. E., et al., 2024, *MNRAS*, 535, 2068
- Sabti N., Muñoz J. B., Kamionkowski M., 2024, *Physical Review Letters*, 132, 061002
- Sampaio V. M., Kolesnikov I., de Carvalho R. R., Ferreras I., Silk J., 2025, *MNRAS*, 541, 1164
- Schaerer D., 2002, *A&A*, 382, 28
- Schaller G., Schaerer D., Meynet G., Maeder A., 1992, *Astronomy and Astrophysics Supplement Series*, 96, 269
- Schaye J., et al., 2015, *MNRAS*, 446, 521
- Schleicher D. R. G., Palla F., Ferrara A., Galli D., Latif M., 2013, *A&A*, 558, A59
- Schmidt W., Federrath C., Hupp M., Kern S., Niemeyer J. C., 2009, *A&A*, 494, 127
- Scholtz J., et al., 2025, Tentative Rotation in a Galaxy at $Z \sim 14$ with ALMA
- Schulz C., Pflamm-Altenburg J., Kroupa P., 2015, *A&A*, 582, A93
- Shakura N. I., Sunyaev R. A., 1973, *A&A*, 24, 337
- Shankar F., Weinberg D. H., Miralda-Escudé J., 2013, *MNRAS*, 428, 421
- Shapiro S. L., Teukolsky S. A., 1986, *Black Holes, White Dwarfs and Neutron Stars: The Physics of Compact Objects*. Wiley-VCH
- Sharda P., Federrath C., Krumholz M. R., 2020, *MNRAS*, 497, 336
- Shi J., et al., 2022, *ApJ*, 927, 189
- Shi Y., Kremer K., Hopkins P. F., 2024b, Feedback-Regulated Seed Black Hole Growth in Star-Forming Molecular Clouds and Galactic Nuclei, doi:10.48550/arXiv.2405.12164
- Shi Y., Kremer K., Hopkins P. F., 2024a, From Seed to Supermassive Black Holes: Capture, Growth, Migration, and Pairing in Dense Proto-Bulge Environments, doi:10.48550/arXiv.2405.17338
- Shi K., Malavasi N., Toshikawa J., Zheng X., 2024c, *ApJ*, 961, 39
- Shuntov M., et al., 2025, Constraints on the Early Universe Star Formation Efficiency from Galaxy Clustering and Halo Modeling of $H\alpha$ and [O III] Emitters, doi:10.48550/arXiv.2503.14280
- Sinha M., Garrison L. H., 2019, in *Software Challenges to Exascale Computing. Second Workshop*. Springer, Singapore, Delhi, India, pp 3–20 (arXiv:1911.08275), doi:10.1007/978-981-13-7729-7_1
- Sinha M., Garrison L. H., 2020, *MNRAS*, 491, 3022
- Sivasankaran A., et al., 2025, *MNRAS*, 537, 817
- Smith B., Sigurdsson S., Abel T., 2008, *MNRAS*, 385, 1443
- Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008a, *MNRAS*, 391, 481
- Somerville R. S., et al., 2008b, *ApJ*, 672, 776
- Spinoso D., Bonoli S., Valiante R., Schneider R., Izquierdo-Villalba D., 2023, *MNRAS*, 518, 4672
- Springel V., 2005, *MNRAS*, 364, 1105
- Springel V., Hernquist L., 2003, *MNRAS*, 339, 289

- Springel V., Hernquist L., 2005, *ApJ*, 622, L9
- Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, *MNRAS*, 328, 726
- Stacy A., Bromm V., Loeb A., 2011, *ApJ*, 730, L1
- Stacy A., Bromm V., Lee A. T., 2016, *MNRAS*, 462, 1307
- Stiskalek R., Desmond H., Devriendt J., Slyz A., Lavaux G., Hudson M. J., Bartlett D. J., Courtois H. M., 2025, *arXiv e-prints*, p. arXiv:2502.00121
- Stone M. A., Lyu J., Rieke G. H., Alberts S., 2023, *ApJ*, 953, 180
- Stone M. A., Lyu J., Rieke G. H., Alberts S., Hainline K. N., 2024, *ApJ*, 964, 90
- Sturm E., et al., 2011, *ApJ*, 733, L16
- Su K.-Y., Bryan G. L., Haiman Z., Somerville R. S., Hayward C. C., Faucher-Giguère C.-A., 2023, *MNRAS*, 520, 4258
- Su K.-Y., Bryan G. L., Haiman Z., 2025, *MNRAS*
- Suess K. A., et al., 2024, *ApJ*, 976, 101
- Sun J., et al., 2025, *ApJ*, 989, L26
- Susa H., 2013, *ApJ*, 773, 185
- Tacchella S., et al., 2023, *ApJ*, 952, 74
- Tinker J. L., 2021, *ApJ*, 923, 154
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, *ApJ*, 688, 709
- Topping M. W., et al., 2024, *MNRAS*, 529, 3301
- Tremaine S. D., Ostriker J. P., Spitzer Jr. L., 1975, *ApJ*, 196, 407
- Tremmel M., Governato F., Volonteri M., Quinn T. R., 2015, *MNRAS*, 451, 1868
- Tsatsi A., Mastrobuono-Battisti A., van de Ven G., Perets H. B., Bianchini P., Neumayer N., 2017, *MNRAS*, 464, 3720
- Übler H., et al., 2023, *A&A*, 677, A145
- Valiante R., Schneider R., Salvadori S., Bianchi S., 2011, *MNRAS*, 416, 1916
- Van Den Bosch F. C., Yang X., Mo H., 2003, *MNRAS*, 340, 771
- Ventura E. M., Qin Y., Balu S., Wyithe J. S. B., 2024, *MNRAS*, 529, 628
- Vink J. S., de Koter A., Lamers H. J. G. L. M., 2001, *A&A*, 369, 574
- Virtanen P., et al., 2020, *Nature Methods*, 17, 261
- Vogelsberger M., Genel S., Sijacki D., Torrey P., Springel V., Hernquist L., 2013, *MNRAS*, 436, 3031
- Voit G. M., Oppenheimer B. D., Bell E. F., Terrazas B., Donahue M., 2024, *ApJ*, 960, 28
- Wagner A. Y., Umemura M., Bicknell G. V., 2013, *The Astrophysical Journal*, 763, L18
- Wang H., Mo H. J., Jing Y. P., Yang X., Wang Y., 2011, *MNRAS*, 413, 1973
- Wang L., Dutton A. A., Stinson G. S., Macciò A. V., Penzo C., Kang X., Keller B. W., Wadsley J., 2015, *MNRAS*, 454, 83
- Wang H., et al., 2016, *ApJ*, 831, 164
- Wang H., et al., 2018, *ApJ*, 852, 31
- Wang K., Mo H. J., Li C., Meng J., Chen Y., 2020, *MNRAS*, 499, 89
- Wang C., et al., 2022, *ApJ*, 936, 95
- Wang K., Peng Y., Chen Y., 2023, *MNRAS*, 523, 1268
- Wang T., et al., 2024a, *Nature*, 632, 1009
- Wang X., et al., 2024b, *ApJ*, 967, L42
- Wang T., et al., 2025, *ApJ*, 988, L35
- Weinberger R., et al., 2017, *MNRAS*, 465, 3291
- Weinberger R., et al., 2018, *MNRAS*, 479, 4056
- Weinberger R., et al., 2023, *MNRAS*, 523, 1104
- Wise J. H., Abel T., 2007, *ApJ*, 671, 1559
- Wise J. H., Regan J. A., O’Shea B. W., Norman M. L., Downes T. P., Xu H., 2019, *Nature*, 566, 85
- Wright A. H., et al., 2016, *MNRAS*, 460, 765
- Wright A. C., Brooks A. M., Tremmel M., Young J. E., Munshi F., Quinn T., 2025, *The Merger-Driven Formation of Classical Low Surface Brightness Galaxies in Romulus25*, doi:10.48550/arXiv.2507.21231
- Xiao M., et al., 2024, *Nature*, 635, 311
- Xing Q.-F., et al., 2023, *Nature*, 618, 712
- Yan Z., Jerabkova T., Kroupa P., 2023, *A&A*, 670, A151
- Yang X., Mo H. J., van den Bosch F. C., 2003, *MNRAS*, 339, 1057
- Yang X., Mo H. J., van den Bosch F. C., Pasquali A., Li C., Barden M., 2007, *ApJ*, 671, 153
- Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, *ApJ*, 752, 41
- Yoon Y., Park C., Chung H., Lane R. R., 2022, *ApJ*, 925, 168
- Yoshida N., Abel T., Hernquist L., Sugiyama N., 2003, *ApJ*, 592, 645
- Yuan F., Narayan R., 2014, *Annual Review of Astronomy and Astrophysics*, 52, 529
- Yue M., et al., 2024a, *ApJ*, 966, 176
- Yue M., Eilers A.-C., Ananna T. T., Panagiotou C., Kara E., Miyaji T., 2024b, *ApJ*, 974, L26
- Yung L. Y. A., Somerville R. S., Finkelstein S. L., Wilkins S. M., Gardner J. P., 2023, *Are the Ultra-High-Redshift Galaxies at $z > 10$ Surprising in the Context of Standard Galaxy Formation Models?*, doi:10.48550/arXiv.2304.04348
- Zhang H., Behroozi P., Volonteri M., Silk J., Fan X., Hopkins P. F., Yang J., Aird J., 2022, *MNRAS*, 518, 2123
- Zhang Z., Chen Y., Rong Y., Wang H., Mo H., Luo X., Li H., 2025a, *Nature*, 642, 47
- Zhang Z., Luo B., Jiang L., Brandt W. N., Huang J., Ni Q., 2025b, *ApJ*, 983, 36
- Zhao D. H., Mo H. J., Jing Y. P., Börner G., 2003, *MNRAS*, 339, 12
- Zhao D. H., Jing Y. P., Mo H. J., Boerner G., 2009, *ApJ*, 707, 354
- Zheng Z., et al., 2005, *ApJ*, 633, 791
- Zhuang M.-Y., Ho L. C., 2023, *Nature Astronomy*, pp 1–14
- Zinger E., et al., 2020, *MNRAS*, 499, 768
- van Dokkelaar F., Agertz O., Renaud F., 2022, *MNRAS*, 512, 3806
- van Dokkelaar F., Mayer L., Capelo P. R., Tamfal T., Quinn T. R., Madau P., 2023, *MNRAS*, 522, 1726

APPENDIX A: DETAILED IMPLEMENTATION OF THE MODEL

In §5.1, we have listed the steps to implement the model. Here we describe the implementation in detail following the logical order outlined in the flowchart in Fig. 8. For the physical contexts for the seeding and growing procedures, refer to Figs. 2 and 5, respectively. The code for the implementation is publicly available (see). For reference, we also list the components of the model and the default parameters adopted in this paper in Table A1.

The model starts with a set of subhalo merger trees constructed from a N-body simulation (TNG100-1-Dark in this paper; see §2). A cosmological simulation covering such a volume often has limited resolution, and is thus not enough to resolve the assembly histories of halos down to the masses required by the seeding procedure. To tackle this issue, we extend the simulated merger trees by joining them with those generated by a Monte Carlo algorithm based on the extended Press-Schechter formalism. The Monte Carlo trees need to be generated with sufficiently high resolution, so that the assembly history of any simulated halo, after the extension, can be reliably traced back to a mass below the H_2 -cooling limit at which the first generation of stars and BH seed may form (Eq. 14; see also the purple dotted lines in Fig. 1a, c and d).

A1 Extension of subhalo merger trees

We follow the algorithm detailed in Chen et al. (2023a, see their §3.2, the ‘Central-Stage Completion’ part) to perform the extension. We list the steps below.

- (i) We decompose each simulated subhalo merger tree into a set of disjoint branches by a recursive method. Starting from the forest F that consists of all the simulated subhalo merger trees, we arbitrarily select a tree T from F . We find the root subhalo, h_{root} , of T , and trace back the main branch, B , of h_{root} (defined as the branch with ‘most massive’ history in TNG; see e.g. Rodriguez-Gomez et al. 2015). We

Table A1. List of model components and default parameters for each component. See §5.1 for a brief description of the implementation and Fig. 8 for a flowchart outlining the components and pipeline. A full description of the implementation, as well as the sources of the default parameters, is presented in Appendix A.

Model Component	Default Parameters	Outline and Implementation
Extension of subhalo merger trees	$N_{\text{MC}} = 512; M_{v,\text{anc}} = 5 \times 10^8 h^{-1} M_{\odot}$	Fig. 8a and Appendix A1
The global integrator	$\sigma_{\text{M}} = 0.2 \text{ dex}$	Appendix A2
The seeding procedure (§3)	Pop-III threshold: $Z_{\text{IGM,Pop-III}} = 10^{-4} Z_{\odot}$; Delay of seeding: $J_{\text{crit},21} = 7.5; \beta_{\text{J}} = 1.5; \gamma_{\text{crit}} = 10; \beta_{\gamma} = 2$; Masses of seeds: $\eta_{\text{r}} = 25; \beta_{\text{m}} = 1; M_{\text{BH,Pop-II}} = 10 M_{\odot}; M_{*,\text{Pop-II}} = 100 M_{\odot}$; $M_{\text{g,seed}} = 10 M_{\odot}$	Fig. 8b and Appendix A3
The growing procedure (§4)	Halo assembly: $\gamma_{\text{f}} = 3/16$; Gas cooling: $M_{\text{cool,f}} = 10^{13} h^{-1} M_{\odot}; \beta_{\text{cool,f}} = 4; M_{\text{cool,s}} = 10^{11.9} h^{-1} M_{\odot}$; $\beta_{\text{cool,s,0}} = 3.6; \gamma_{\text{cool,s}} = -0.72$; Feedback (continuous mode): $\alpha_{\text{SN,f}} = 0; \beta_{\text{SN,f}} = 2.5; V_{\text{w}} = 250 \text{ km s}^{-1}$; $R_{\text{s}} = 10^{-0.96}; \beta_{\text{SN,s}} = 1.9; \alpha_{\text{AGN,f,0}} = 5 \times 10^{-3}; \beta_{\text{AGN,f}} = 2; f_{\text{ej,f}} = 0.75$; $f_{\text{ej,s}} = 1$; Star formation and BH accretion (continuous mode): $\epsilon_{*,\text{f}} = 0.75; \epsilon_{*,\text{s}} = 0.32$; $R = 0.4; \alpha_{\text{cap,f}} = 2.5; \alpha_{\text{cap,s}} = 0$; Satellite evolution: $\tau_{\text{sat}} = 4 \text{ Gyr}/\ln 10$; SNF nucleus: $\delta_{\gamma} = 1; n_{\text{snf}} = 10^{3.5} \text{ cm}^{-3}; p = 0; q = 0.5$; BH accretion disk: $\zeta_{\text{t}} = \zeta_{\text{A}} = 1/3; \psi_{\text{t}}/\psi_{\text{A}} = 1$; Nuclear burst: $\xi_{\text{AGN},\infty} = 1; v_{\text{out}} = 10^3 \text{ km s}^{-1}; \beta_{\text{out}} = 0; \alpha_{\text{SF}} = 1; \alpha_{\text{BH}} = 1$; $\delta_{\text{Q}} = 1; f_{\text{dry}} = 6.8 \times 10^{-3}$; Merger and dynamical heating: $f_{\text{merge}} = 0.5; f_{\text{heat}} = 0.5$; Metallicity: $V_{\text{esc}} = 75 \text{ km s}^{-1}; \beta_{\text{esc}} = 2; z_{\text{esc}} = 5; \gamma_0 = 1; \beta_{\text{mix}} = 2$	Fig. 8c and Appendix A4
The model of LW radiation field	$f_{\text{esc}} = 1; \eta_{\text{LW}} = 4 \times 10^3; Q_{\text{LW}} = 10^{36} \text{ erg} \cdot \text{s}^{-1} M_{\odot}^{-1}; J_{\text{local,th},21} = 0.01$	Fig. 8c2 and Appendix A5
The model of IGM enrichment	$A_{\text{sh}} = 1; \epsilon_{\text{SN}} = 10^{51} \text{ erg} \cdot M_{\odot}^{-1}; l_{\text{sm}} = 250 h^{-1} \text{ kpc}$;	Fig. 8c2 and Appendix A6

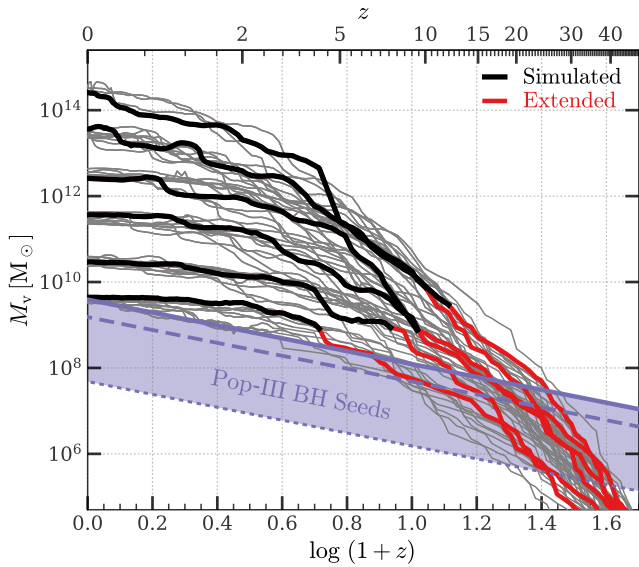


Figure A1. Extension of subhalo merger trees. Here we show the main-branch mass assembly histories (MAHs) for examples of central subhalos at $z = 0$ with masses separated by 1 dex in the range of $10^{9.5} h^{-1} M_{\odot} - 10^{14.5} M_{\odot}$. For each curve, **black** part is obtained directly from the simulation, which is truncated at the anchor point due to limited resolution; **red** part shows the extension performed by our Monte Carlo algorithm that smoothly join the simulated part at the anchor point. Each bundle of **grey** curves show the MAHs of up to 10 additional halos with masses at $z = 0$ similar to the example. The extension ensures that the assembly histories can be traced back to sufficiently high redshift and sufficiently low halo mass so that the formation of Pop-III stars and BH seeds can be captured, as indicated by the **purple** band (the same as that in Fig. 1). See Appendix A1 for details.

remove B from T , which leaves T with a set of disjoint subtrees, $\{T_i\}$, each rooted in a subhalo h_i whose descendant is h_{root} . We put all T_i back into F , and repeat the above process until F is empty.

- (ii) For each branch B obtained above, we start from the leaf subhalo (i.e. that with the highest redshift), walk along the branch towards lower redshift to find the ‘anchor’ redshift z_{anc} at which the halo mass M_v becomes larger than a threshold $M_{v,\text{anc}}$ for the first time. The subhalo at z_{anc} so obtained is referred to as the ‘anchor subhalo’ of B , and is denoted as h_{anc} .
- (iii) We start from h_{anc} , walk along the branch B towards higher redshift to find the ‘half-mass’ redshift, z_{form} , of h_{anc} , at which the halo mass drops below $M_v(z_{\text{anc}})/2$ for the first time.
- (iv) We follow Parkinson et al. (2008) to generate a set of N_{MC} Monte Carlo merger trees with a sufficiently high resolution, each having a root halo h'_{root} with a mass and redshift equal to that of h_{anc} . We compute the half-mass formation time, z'_{form} , for the root halo of each Monte Carlo tree along the main branch.
- (v) We search for the Monte Carlo tree whose root halo has a z'_{form} that is the closest to z_{form} among all the Monte Carlo trees, and we denote the main branch of this root halo as B' . Physical quantities directly related M_v (e.g. R_v , V_v and T_v ; see Eq. 3–5) are assigned to the halos of B' according to their definitions. The profile of any halo in B' is assumed to be an NFW profile, with a concentration parameter (c_v ; see Eq. 6) computed from the mass assembly history by the analytical fitting of Zhao et al. (2009, with slight adaptation to account for the difference in the definition of halo boundary), and the maximal circular velocity (V_{max}) is computed from the profile.
- (vi) We prune all halos at $z > z_{\text{anc}}$ from B , and join B' to B at z_{anc} .

Note that the above steps ensure that the Monte Carlo branch B' not only has exactly the same M_v as B at z_{anc} , but also has similar the first-order derivative, since the matching of z_{form} approximately

provides such a constraint. The Monte Carlo algorithm proposed by Parkinson et al. (2008) is adopted here. This algorithm is re-calibrated by N-body simulations to closely match the simulation results (see Jiang & van den Bosch 2014 for details), and is highly efficient due to their carefully chosen bound function for the rejection sampling. The number of Monte Carlo trees, N_{MC} , should be sufficiently large so that the sampling of z'_{form} is dense enough to give a close match to z_{form} . By trial and error, we find that $N_{\text{MC}} = 512$ is sufficient to produce a smooth extension. The threshold mass $M_{\text{v,anc}}$ that defines the anchor point should be at least twice the minimal mass that can be resolved by the simulation used, so that z_{form} of h_{anc} can be reliably determined. Based on the resolution of TNG100-1-Dark, we set $M_{\text{v,anc}} = 5 \times 10^8 h^{-1} M_{\odot}$, which is the mass of about 100 dark-matter particles in the simulation. Since the seeding procedure requires modeling the radiation and matter feedback from any galaxy to its large-scale environment (see the flowchart Fig. 8c2), the locations of halos in each B' have to be assigned. Here, we assign the location of h_{anc} in B to all halos in the matched B' . This means that the peculiar motion of a halo at $z > z_{\text{anc}}$ is ignored and that the accuracy of small-scale radiation and matter feedback is compromised.

With the extension, the assembly history of each simulated subhalo can be smoothly extended down to the H_2 -cooling limit, and the extended merger trees will be used as the input of subsequent modeling for galaxy formation. The steps for obtaining the simulated merger trees and extending them are outlined in the flowchart of Fig. 8a.

In Fig. A1, we show examples of mass assembly histories (MAHs, i.e. M_{v} at different redshifts) of simulated subhalos before (black) and after (black plus red) the extension, over-plotted with the regime in which Pop-III stars are expected to form (purple). As expected, the MAHs now cross the Pop-III band, thus allowing the formation of the first generation of stars and BH seeds to be directly modeled. Note that the Monte Carlo trees themselves can be used as the input of the model, if LW radiation and IGM enrichment are available by other means. This is the case for the idealized and controlled experiments presented in §5.3.3, where the environment is set to the extreme so that $M_{\text{BH,seed}}$ are minimized or maximized. A similar approach was used by Spinoso et al. (2023, see their §2.4 and fig. 1) to marginally extend the simulated merger trees to the H_2 cooling limit at $z \gtrsim 15$, thereby allowing some light-weight seeds to be reared. The semi-analytical model of Ventura et al. (2024) for Pop-III stars was built on merger trees from a N-body simulation that can resolve halos down to $\approx 1.5 \times 10^5 h^{-1} M_{\odot}$, but with a compromised volume of $(10 h^{-1} \text{Mpc})^3$. Our mixed approach of using a large-volume simulation and a Monte Carlo extension thus allows us not only to trace the formation of Pop-III stars, but also to have a large sample that can reduce the effects of cosmic variance (e.g. Chen et al. 2023b; Jespersen et al. 2025).

A2 The global integrator

The seeding procedure in our model is built on two factors in the large-scale environment of subhalos: the LW radiation and the IGM metallicity. Both factors are self-consistently modeled by tracing the radiation and matter feedback from all other subhalos that have already formed stars by the time step in question. Entanglement among subhalo merger trees is thus inevitable and the integration of the formation histories of galaxies must be performed globally, rather than individually on each subhalo merger tree. Here we refer to this integration process as the ‘global integrator’, and we implement it as follows.

- (i) Starting from the snapshot with the highest z , we iterate over all snapshots towards the lower z until $z = 0$.
- (ii) For each snapshot s (at redshift z_s), we take all the N_s subhalos, $H_s \equiv \{h_i\}_{i=1}^{N_s}$, belonging to s from the extended merger trees constructed in Appendix A1. For each subhalo $h \in H_s$, we follow the algorithms to be detailed in Appendices A5 and A6, respectively, to find the LW intensity (J_{LW}) and the IGM metallicity (Z_{IGM}) at \mathbf{x}_h , the location of h . This step is schematically shown in Fig. 8c2 as a feedback loop between the two main procedures (seeding and growing).
- (iii) We check whether h has already formed stars at $z > z_s$. If so, we assume the galaxy in h has been self-enriched, and branch to the growing procedure to be detailed in Appendix A4 (see also Fig. 8c for an outline) to model the post-seeding growth of the galaxy. If not, we follow the seeding procedure to be detailed in Appendix A3 (see also Fig. 8b, including the inset c1, for an outline) to check whether the condition for the collapse of gas in h is satisfied, and breed a BH seed if the condition is satisfied. Otherwise, we assume that h remains ‘dark’, without any star or BH formed at z_s .

With this global integrator, each subhalo is assigned with a seeding epoch at which the first generation of stars and BH seed is bred, and the post-seeding growth is driven by different channels that gradually build up the BH, stellar and gas contents of the subhalo. As demonstrated by Chen et al. (2023b), both physical and observational uncertainties in the predicted quantities can introduce biases in statistical properties of galaxy populations, especially for rare objects. Accurately accounting for such uncertainties is almost impossible, and here we provide a tentative solution by injecting stochastic noise into the predicted quantities. As shown in e.g. Chen et al. (2021a) and Chen et al. (2021b) using hydro simulations and regressors with successively increased number of predictors, many nuanced factors other than halo properties can produce galaxy-to-galaxy variance in the growth, and such variance may thus be missed by a halo-based model. To account for such missed variance, we introduce a stochastic noise, with a log-normal distribution that has a zero mean and a standard deviation of $\sigma_{\text{M}} = 0.2$ dex, to the modeled amount of growth, ΔM_{BH} and $\Delta M_{\text{*}}$, in every time step, at every growth stage (fast or slow phases, or satellite stage; see below) and through every growth channel (seeding, or bursty or continuous modes). In comparison with observations, when appropriate, we additionally add stochastic noise to mimic observational uncertainties, which will be clearly stated when the comparison is made (see Fig. 17 and §5.4).

A3 The seeding procedure

The seeding procedure takes the internal and environmental properties of a subhalo $h \in H_s$ that has not yet formed any star as input, determines whether or not the first generation of stars and the BH seed can form in h at z_s , and breeds stars and BH seed. The procedure follows prescriptions presented in §3, and is implemented through the following steps (see Fig. 8b for an outline).

- (i) We check the IGM metallicity, Z_{IGM} , at \mathbf{x}_h . If $Z_{\text{IGM}} \geq Z_{\text{IGM,Pop-III}}$, we populate h with a Pop-II star cluster, breed a BH seed from the cluster, and end the seeding procedure for h . The mass of the BH seed is set to be $M_{\text{BH,Pop-II}}$, and the mass of the remaining stars in total is set to be $M_{\text{*},\text{Pop-II}}$.
- (ii) Otherwise ($Z_{\text{IGM}} < Z_{\text{IGM,Pop-III}}$), the subhalo is in the Pop-III regime. We take the LW intensity (J_{LW}) at \mathbf{x}_h and the specific accretion rate (γ_{v}) of h , and use Eq. (23) to find the threshold of halo mass ($M_{\text{v,th}}$) for the gas in h to collapse. If the mass of h is below the threshold ($M_{\text{v}} < M_{\text{v,th}}$), we assume that h forms no star nor BH at z_s , and

we end the seeding procedure for h . Otherwise ($M_v \geq M_{v,\text{th}}$), we assume that a Pop-III star cluster forms in h , and proceed to the next step to breed a BH seed from the cluster.

- (iii) To determine the properties of the Pop-III star cluster, we first evaluate the properties of the SGC in h according to Eqs. (9)–(13), assuming that a pristine halo has a baryon fraction equal to the cosmic value, f_b (see texts around Eq. 1). We then use the radiation-regulated scenario (§3.2) to determine the mass of the dominant star in the cluster. Specifically, we use Eq. (28) to compute the zero-age main sequence star formation efficiency (ϵ_{ZAMS}) and luminosity (L_{ZAMS}) of the dominant star. If this luminosity does not exceed the Eddington luminosity ($L_{\text{ZAMS}} < L_{\text{Edd}}$), the mass of the dominant star is set to be $M_{*,1\text{st}} = \epsilon_{\text{ZAMS}} M_{\text{SGC}}$. Otherwise ($L_{\text{ZAMS}} \geq L_{\text{Edd}}$), the star reaches the Eddington-level accretion, and we turn to Eq. (31) to find the corresponding star formation efficiency (ϵ_{Edd}) and set $M_{*,1\text{st}} = \epsilon_{\text{Edd}} M_{\text{SGC}}$.
- (iv) With the mass of the dominant star, the IMF of the entire Pop-III cluster is obtained by Eq. (34), which is a power-law function with an index of $-\beta_m$ and an exponential cutoff at $M_{*,1\text{st}}$. The normalization of the IMF is obtained by solving Eq. (35) that equates the integral of the IMF over $[M_{*,1\text{st}}, +\infty]$ to 1. This implies that the instantiation of stars from the IMF is a process of ‘optimal sampling’, which is supported by observations of star clusters (Yan et al. 2023), and is also consistent with theoretical expectation under our assumption of self-regulation (Kroupa et al. 2013; Schulz et al. 2015).
- (v) The evolution of the dominant star is treated by a simplified piecewise model (see §3.2), with two mass thresholds, $M_{*,\text{PISN}}$ and $M_{*,\text{DCBH}}$. If $M_{*,1\text{st}} \geq M_{*,\text{DCBH}}$, the star ends its life as a DCBH, which we use as the BH seed. The mass of the seed is set to be $M_{\text{BH,seed}} = M_{*,1\text{st}}$, assuming no significant mass loss during the stellar evolution and the direct collapse. If $M_{*,\text{PISN}} \leq M_{*,1\text{st}} < M_{*,\text{DCBH}}$, the star ends its life by PISN which totally disrupts the star and leaves no remnant. In this case, we seek for the most massive star (with a mass $M_{*,\text{surv}}$) that can survive PISN by requiring the integral of the IMF over $[M_{*,\text{surv}}, M_{*,\text{PISN}}]$ to be equal to 1 (Eq. 38), again implying optimal sampling. The remnant of this star is treated as the BH seed, and the mass of the seed is set to be $M_{\text{BH,seed}} = f_{\text{rem}} M_{*,\text{surv}}$, accounting for a fraction of $1 - f_{\text{rem}}$ of mass loss during the stellar evolution. If $M_{*,1\text{st}} < M_{*,\text{PISN}}$, the dominant star ends its life by CCSN, and leaves a BH remnant with a mass of $M_{\text{BH,seed}} = f_{\text{rem}} M_{*,1\text{st}}$, which is adopted as the BH seed. Once the BH seed is determined, the evolution and remnant mass of any other star in the cluster is obtained similarly by the piecewise model, weighted by the IMF, and integrated over $[M_{*,\text{min}}, M_{*,1\text{st}}]$ (for DCBH or CCSN) or $[M_{*,\text{min}}, M_{*,\text{surv}}]$ (for PISN) to obtain $M_{*,\text{seed}}$, the stellar mass bred together with the BH seed.

The default value of parameters are chosen as follows. The critical metallicity is set to be $Z_{\text{IGM,Pop-III}} = 10^{-4} Z_{\odot}$ (see §3.2). The parameters describing the dependence of the collapse mass on LW radiation are set to be $J_{\text{crit},21} = 7.5$ and $\beta_J = 1.5$ (see §3.1.3), and on dynamical heating are set to be $\gamma_{\text{crit}} = 10$ and $\beta_{\gamma} = 2$ (see §3.1.4). The coupling efficiency in the radiation-regulated scenario is set to be $\eta_r = 25$ (see §3.2). The index of the IMF of Pop-III stars is set to be $\beta_m = 1$ (see §3.2). The mass of a Pop-II BH seed is set to be $M_{\text{BH,Pop-II}} = 10 M_{\odot}$, a typical value for the remnant of a massive star in normal star clusters. Unfortunately, the stellar mass associated with a Pop-II BH seed is unknown, and so an assumption has to be made. We notice that the bursty mode dominates the growth of galaxies after the seeding, erases the difference produced by the seeding method, and sustains a ratio of $M_{\text{BH}}/M_* \approx 1/10$ (see §5.3.1, Figs. 14 and 17). Thus, we manually set $M_{*,\text{Pop-II}} = 10 M_{\text{BH,Pop-II}}$, so that a Pop-

II seed appears in a $M_{\text{BH}}-M_*$ relation similar to that shaped by the post-seeding growth without introducing an artificial discontinuity. In both Pop-II and Pop-III cases, the gas density at fragmentation is expected to be high due to the lack of effective metal cooling, and thus the seeding process may be bursty. This implies a strong gas depletion during seeding, and a short period of quiescent phase after seeding, as discussed in the beginning of §4. For this reason, we set the remaining gas mass after seeding to be negligibly small, $M_{\text{g,seed}} = 10 M_{\odot}$.

With the seeding procedure, the halo h either remains ‘dark’, or is populated with initial values of $M_{\text{BH,seed}}$, $M_{*,\text{seed}}$ and $M_{\text{g,seed}}$ so that the post-seeding processes modeled by the growing procedure are provided with proper initial conditions.

A4 The growing procedure

The growing procedure takes a subhalo $h \in H_s$ that has already been seeded as the input, and updates the properties of the galaxy in h by modeling the evolution within the current time step (i.e. the duration since the previous snapshot, $s - 1$, to the current snapshot s). The growing procedure is based on the last version of our model presented in §2 of Paper-III, with a major extension to include the growth in the bursty mode proposed in this paper and a few updates to accommodate the extension on the other components of the model. Unless otherwise specified, the default parameters adopted in this paper are taken from Paper-III.

In terms of the underpinning of BH growth, the assembly of dark matter halos (quantified by γ_v ; see Eq. 19) is expected to be the key factor that determines the gas dynamics of the central galaxy hosted by the halo, and thus also determines how the central BH hosted by the galaxy accretes the gas and reacts via AGN feedback (see the discussion in §4.1 and summary in Fig. 5). Specifically, the BH in a halo during the fast phase grows efficiently by capturing the subclouds formed out of the dynamically hot (turbulent) SGC, while the BH in a halo during the slow phase grows inefficiently due to the barrier of angular momentum (j) associated with the formation of dynamically cold disk. In addition, a global disturbance, quantified in this paper by a temporary excursion of the halo assembly to high γ_v , induces violent gas inflow that causes the formation of an SNF nucleus, followed by a nuclear burst that increases the mass of the BH.

The above physical considerations motivate us to treat galaxy formation as a multichannel process (see §5.3.1, Figs. 13 and 14). To aid the implementation of the model, we separate each branch B of subhalo merger trees constructed in Appendix A1 into several stages/phases, so that numerical prescriptions for modeling galaxies hosted by subhalos during each stage can be applied separately. For this, we first identify the ‘infall snapshot’ (s_{infall}) of B , defined as the last snapshot in which the subhalo in B is still a central subhalo. Subhalos at and before s_{infall} are defined to be in the ‘central stage’, and those after s_{infall} are defined to be in the ‘satellite stage’. For numerical convenience, we set the value of M_v for any subhalo in the satellite stage to be that at s_{infall} , effectively assuming that there is no growth or excursion of halo assembly in the satellite stage. For the entire branch, we smooth the MAH (i.e. values of M_v) by a running Gaussian kernel with a standard deviation equal to the dynamical timescale of the halo (t_v ; see Eq. 7), and denote the smoothed MAH as $\mathcal{S}(M_v; t_v)$. We then perform a numerical differentiation to obtain the smoothed version of the specific accretion rate from $\mathcal{S}(M_v; t_v)$, and denote it as $\mathcal{S}(\gamma_v; t_v)$. If $\mathcal{S}(\gamma_v; t_v) \geq \gamma_f$, where the default value of the threshold is $\gamma_f = 3/16$, the subhalo is defined to be in the fast phase; otherwise, it is defined to be in the slow phase. Note that this

separation of fast and slow phases is slightly different from the one with a single transition redshift z_f used in previous papers of this series: the excursion of the halo assembly to the slow (fast) phase on a timescale of t_v can now take place at higher (lower) z , thus allowing the model to capture some recurrent transition of galaxy dynamics that has been evident in observations (Parlanti et al. 2023; Rowland et al. 2024; Scholtz et al. 2025; Danhaive et al. 2025; see also the discussion in Appendix C of Paper-I). For convenience, we use B_{fast} , B_{slow} and B_{sat} to denote the sets of subhalos in B during the fast phase, slow phase and satellite stage, respectively.

During the current time step, we determine whether the subhalo h has a temporary excursion to high γ_v by examining the condition $\gamma_v - S(\gamma_v; t_v) \geq \delta_\gamma \gamma_f$, with the default value of $\delta_\gamma = 1$ (see Eq. 42). If the condition is not satisfied, we assume that the growth in the bursty mode is not triggered during this time step. Otherwise, we assume that a global disturbance has been imparted to the SGC, leading to the formation of an SNF nucleus in the SGC center, followed by a nuclear burst that depletes the gas from the nucleus. Note that such a burst takes place in a much shorter timescale (\lesssim Myr; see Fig. 7) and within a much smaller volume (see Fig. 6) than those modeled using the global properties for the host galaxy. To properly treat this small-scale process, a refinement engine is designed and applied on top of the large-scale modeling (see §4.5). Another example of small-scale processes is the formation of individual sub-clouds and star clusters, which is modeled using a similar refined treatment, as detailed in Paper-III. Since these small-scale processes are critical in determining the growth of compact objects, such as GCs, NSCs and SMBHs, which are of particular interests in recent observations at high z , we represent them as a separate inset c1 in Fig. 8.

Other unresolved processes, such as stellar evolution and baryon cycling in individual galaxies, are treated by instantaneous approximation given that their timescales are short compared with the assembly of halos. Hence, the stellar mass formed in any galaxy during any time step, ΔM_* , should be understood as the remaining mass of stars after stellar evolution. On the other hand, if instantaneous (unevolved) star formation rate (SFR) is required, we obtain it by

$$\text{SFR} = \frac{1}{1-R} \frac{\Delta M_*}{\Delta t}, \quad (\text{A1})$$

where $R = 0.4$ is the returned fraction of mass assuming a Chabrier IMF (Bruzual & Charlot 2003) for stellar populations other than Pop-III. Note that in Fig. 15, we have used this unevolved SFR to produce the cosmic star formation history.

With the above definitions, the steps of the growing procedure for a subhalo $h \in H_s$ are as follows (see also Fig. 8c for an outline).

- (i) We find the branch B that contains h , and determine which stage/phase h is in ($h \in B_{\text{fast}}$, $h \in B_{\text{slow}}$ or $h \in B_{\text{sat}}$). If $h \in B_{\text{sat}}$, we jump to step (ix); otherwise, we continue with the next step to model the growth of the BH with its host galaxy and subhalo in the continuous mode.
- (ii) We compute the total amount of gas in the h that is available for subsequent processes:

$$\Delta M_{g,\text{avail}} = \begin{cases} f_b \Delta M_v, & \text{if } h \in B_{\text{fast}}; \\ \frac{f_b M_v}{\tau(z)} \Delta t, & \text{if } h \in B_{\text{slow}}, \end{cases} \quad (\text{A2})$$

where $\tau(z) = (1+z)^{-3/2}/(10H_0)$ is an approximation to the dynamical timescale of the halo (Eq. 7), Δt is the duration of the current time step, and ΔM_v is the change of halo mass in this duration. Note that the equation in the second line is expressed by M_v , instead of ΔM_v , which takes into account that halo growth ceases in the slow phase, and the gas processed earlier and retained in the halo is recycled back

to fuel star formation and BH growth (see e.g. §5.3.2 and Fig. 15 for a discussion).

- (iii) The fraction of available gas in the h that can cool down is determined by a cooling factor, F_{cool} , which is

$$F_{\text{cool}} = \begin{cases} \left[1 + \left(\frac{M_v}{M_{\text{cool},f}} \right)^{\beta_{\text{cool},f}} \right]^{-1}, & \text{if } h \in B_{\text{fast}}; \\ \left(1 + \frac{M_v}{M_{\text{cool},s}} \right)^{-\beta_{\text{cool},s}(z)}, & \text{if } h \in B_{\text{slow}}. \end{cases} \quad (\text{A3})$$

The factor takes into account the fact that cooling is effective only in halos with mass smaller than a threshold of $\sim 10^{12} M_\odot$, and that only a fraction of gas in halos with masses larger than the threshold can cool down and flow into the galaxy (see e.g. the cooling curves in fig. 2 of Paper-I). The default parameters are calibrated to be $M_{\text{cool},f} = 10^{13} h^{-1} M_\odot$, $\beta_{\text{cool},f} = 4$, $M_{\text{cool},s} = 10^{11.9} h^{-1} M_\odot$, $\beta_{\text{cool},s} = \beta_{\text{cool},s,0}(1+z)^{\gamma_{\text{cool},s}}$, $\beta_{\text{cool},s,0} = 3.6$ and $\gamma_{\text{cool},s} = -0.72$. Note that the redshift dependence of both M_v and $\beta_{\text{cool},s}$ makes F_{cool} a function of redshift. With the cooling factor, the amount of gas that can cool down and flow into the galaxy is

$$\Delta M_{g,\text{cool}} = F_{\text{cool}} \Delta M_{g,\text{avail}}, \quad (\text{A4})$$

and the remaining fraction, $1 - F_{\text{cool}}$ is referred to as the ‘hot’ gas. Note that the reduction of the specific growth rate of gas mass in the massive example at low z in Fig. 14c8 is partly due to the reduction of F_{cool} .

- (iv) Owing to feedback effects from star formation and AGN, only a fraction of the cooled gas can cool further and become star-forming gas. We model the amount of star-forming gas by modulating the amount of cooled gas $\Delta M_{g,\text{cool}}$ with two additional factors describing the two sources of feedback, as:

$$\Delta M_{g,\text{SF}} = F_{\text{SN}} F_{\text{AGN}} \Delta M_{g,\text{cool}}. \quad (\text{A5})$$

Here, F_{SN} describes the effect of supernova feedback and is modeled as

$$F_{\text{SN}} = \begin{cases} \frac{\alpha_{\text{SN},f} + (V_g/V_w)^{\beta_{\text{SN},f}}}{1 + (V_g/V_w)^{\beta_{\text{SN},f}}}, & \text{if } h \in B_{\text{fast}}, \\ \left(\frac{M_v/M_{\text{cool},s}}{R_s + M_v/M_{\text{cool},s}} \right)^{\beta_{\text{SN},s}}, & \text{if } h \in B_{\text{slow}}, \end{cases} \quad (\text{A6})$$

where the default parameters are set to be $\alpha_{\text{SN},f} = 0$, $\beta_{\text{SN},f} = 2.5$, $V_w = 250 \text{ km s}^{-1}$, $R_s = 10^{-0.96}$, $\beta_{\text{SN},s} = 1.9$ (Paper-I), and $V_g = V_{\text{max}}$ is the typical velocity dispersion of the galaxy. F_{AGN} describes the effect of AGN feedback, and is given by

$$F_{\text{AGN}} = \begin{cases} \exp \left[-\frac{\alpha_{\text{AGN},f}(z) M_{\text{BH}} c^2}{M_b V_g^2} \right], & \text{if } h \in B_{\text{fast}}; \\ \left(\frac{R_s + M_v/M_{\text{cool},s}}{1 + M_v/M_{\text{cool},s}} \right)^{\beta_{\text{AGN},s}}, & \text{if } h \in B_{\text{slow}}. \end{cases} \quad (\text{A7})$$

Here, M_b is the total baryon mass (gas, stars and BH) of the bulge of the galaxy, and the default parameter $\beta_{\text{AGN},s} = 1.8$. The exponential function for the fast phase is similar to the one adopted in Paper-I at small M_{BH} , but approaches to zero more smoothly at large M_{BH} . Different from previous papers of this series, the efficiency of AGN feedback, as quantified by $\alpha_{\text{AGN},f}(z)$, is now modeled to be redshift dependent. This is partly motivated by the theoretical consideration that the fraction of feedback energy that can be dissipated depends on the cooling timescale of the gas affected by the feedback, and thus also depends on z , owing to the cosmic evolution of the SGC density (Eq. 10). This is also supported by empirical evidences that

star formation efficiency in high mass halos is redshift dependent (viii) (e.g. Lu et al. 2014; Behroozi et al. 2020). Hence, we parameterize the factor $\alpha_{\text{AGN},f}(z)$ as

$$\alpha_{\text{AGN},f}(z) = \alpha_{\text{AGN},f,0}(1+z)^{-\beta_{\text{AGN},f}}. \quad (\text{A8})$$

The value of $\beta_{\text{AGN},f}$ is not known, but the bound of it may be obtainable from a timescale argument that takes into account the competition of energy injection and dissipation. If we consider an extreme case that the accretion of BH smoothly tracks the growth of the host SGC and thus the timescales of BH growth and injection of feedback energy follow the free-fall timescale of the SGC ($t_{\text{ff},\text{SGC}}$; Eq. 13), the feedback efficiency may be approximated as $\alpha_{\text{AGN},f} \sim t_{\text{cool,SGC}}/t_{\text{ff,SGC}} \propto n_{\text{SGC}}^{-1/2}$, where the cooling timescale $t_{\text{cool,SGC}} \propto n_{\text{SGC}}^{-1}$ holds for any two-body radiative process (see e.g. Appendix B.1.4 of Mo et al. 2010). In the other extreme, if the accretion of BH is a highly localized process with a microscopic timescale irrespective of the SGC density, the feedback efficiency may be approximated as $\alpha_{\text{AGN},f} \propto t_{\text{cool,SGC}} \propto n_{\text{SGC}}^{-1}$. The above timescale argument suggests a bound for the feedback efficiency as $\alpha_{\text{AGN},f} \propto n_{\text{SGC}}^{[-1, -1/2]}$, and a more realistic value may be somewhere in between. In this paper, we adopt the default as $\alpha_{\text{AGN},f} \propto n_{\text{SGC}}^{-2/3}$ (i.e. $\beta_{\text{AGN},f} = 2$; see Eq. 10 for n_{SGC}), which falls between the two extremes. The default value of the normalization factor is set to be $\alpha_{\text{AGN},f,0} = 5 \times 10^{-3}$, so that $\alpha_{\text{AGN},f}$ at $z \approx 2$ (the redshift at which massive halos with $M_{\text{v}} \gtrsim 10^{13} M_{\odot}$ ends its fast phase; see Fig. 1b) is comparable to the value calibrated in Paper-I.

- (v) Part of the feedback-affected gas is expected to be expelled from the galaxy (referred to as ‘ejected gas’), while the rest remains hot in the galaxy and is prevented from forming stars (referred to as ‘prevented gas’). To separate the two components, we introduce a parameter f_{ej} , which is the fraction of the feedback-affected gas that is ejected from the galaxy. Thus, the amounts of ejected and prevented gas are

$$\Delta M_{\text{g,ej}} = f_{\text{ej}}(1 - F_{\text{SN}}F_{\text{AGN}})\Delta M_{\text{g,cool}}, \quad (\text{A9})$$

and

$$\Delta M_{\text{g,prev}} = (1 - f_{\text{ej}})(1 - F_{\text{SN}}F_{\text{AGN}})\Delta M_{\text{g,cool}}, \quad (\text{A10})$$

respectively. Following Paper-III, we set the default value to be $f_{\text{ej}} = f_{\text{ej},f} = 0.75$ for the fast phase, so that a galaxy well regulated by feedback has a gas fraction comparable to the typical spin parameter, 0.04, of dark matter halos, and thus can transit to a stable disk in the slow phase. For the slow phase, we set $f_{\text{ej}} = f_{\text{ej},s} = 1$, meaning that all feedback-affected gas is ejected from the disk, with some recycled back later on a dynamical timescale of the host halo (see Eq. A2).

- (vi) The amount of stars formed in the galaxy is related to the amount of star-forming gas, $\Delta M_{\text{g,sf}}$, by a star formation efficiency, ϵ_* , as

$$\Delta M_* = \epsilon_* \Delta M_{\text{g,sf}}, \quad (\text{A11})$$

where $\epsilon_* = \epsilon_{*,f} = 0.75$ and $\epsilon_* = \epsilon_{*,s} = 0.32$ for the fast and slow phases, respectively, as calibrated in Paper-I. The stellar mass so obtained is added to the bulge component of the galaxy if $h \in B_{\text{fast}}$, or to the disk component if $h \in B_{\text{slow}}$.

- (vii) The growth of the mass of the central BH is obtained by the fraction of low- j sub-clouds that can approach the center of the galaxy and be captured by the BH, which is derived in Paper-I as

$$\Delta M_{\text{BH}} = \alpha_{\text{cap}} \frac{M_{\text{BH}}}{M_{\text{b}}} \Delta M_{\text{g,sf}} \quad (\text{A12})$$

where $\alpha_{\text{cap}} = \alpha_{\text{cap},f} = 2.5$ in the fast phase, and $\alpha_{\text{cap}} = \alpha_{\text{cap},s} = 0$ in the slow phase due to the full mixing of gas. Note that the ‘enhancement factor’ is now removed from the equation, as it is replaced explicitly by the nuclear burst (see steps x and xi below).

(viii) The change of gas mass in the galaxy can be obtained by mass conservation, taking into account the contributions by cooling, consumption by star formation and BH accretion, and ejection by feedback, as

$$\Delta M_{\text{g}} = \Delta M_{\text{g,SF}} - \Delta M_* - \Delta M_{\text{BH}} + \Delta M_{\text{g,prev}}. \quad (\text{A13})$$

Similar to the treatment for M_* , the update of gas mass is performed on the bulge and disk components, if $h \in B_{\text{fast}}$ and $h \in B_{\text{slow}}$, respectively. Note that the mass and specific rate of gas in Fig. 14 (rows 7 and 8) are obtained with this equation.

- (ix) If the subhalo is in the satellite stage ($h \in B_{\text{sat}}$), the star formation efficiency is expected to be low due to the environmental effects, such as ram-pressure stripping (Gunn & Gott 1972; Du et al. 2019; Kulier et al. 2023; Carnall et al. 2023) and tidal stripping (Binney & Tremaine 2008; Read et al. 2006; Drakos et al. 2022). Motivated by the observational results of e.g. Peng et al. (2015), we model the SFR in this stage by an exponential decay, as

$$\dot{M}(t) = \dot{M}(t_{\text{infall}}) \exp\left(-\frac{t - t_{\text{infall}}}{\tau_{\text{sat}}}\right), \quad (\text{A14})$$

where $t_{\text{infall}} \equiv t(s_{\text{infall}})$ is the cosmic time at the infall snapshot, and $\tau_{\text{sat}} = 4 \text{ Gyr}/\ln 10$ is the timescale of the decay. The growth of BH is assumed to be stopped in this stage. The amount of gas is updated according to the ‘close-box’ model with instantaneous recycling: the supply of gas is stopped by the environmental effects, and the feedback-affected gas is instantaneously recycled into the ISM. With these assumptions, the only sink term of gas is the consumption of star formation.

- (x) If a nuclear burst is triggered within the current time step, we take M_{g} of bulge (obtained from Eq. A13) as the mass of SGC (M_{SGC}), derive the radius fraction, $f_{\text{r}} \equiv R_{\text{SGC}}/R_{\text{v}}$, by the self-gravitating condition, as $f_{\text{r}} = M_{\text{SGC}}/M_{\text{v}}$ (see Eq. 26 of Paper-I), and use Eqs. (9)–(13) to compute the properties of the SGC. Using the properties of the SGC, we move a step inward to the galactic nucleus according to double power-law density profile of the entire SGC (Eq. 44; see also Fig. 6 for examples), and compute the properties of the SNF nucleus using Eqs. (45)–(48). The default parameters to define the SNF nucleus include the supernova-free density, $n_{\text{snf}} = 10^{3.5} \text{ cm}^{-3}$, and the structural parameters, $p = 0$ and $q = 0.5$. Moving a step further into the zone of influence of the BH ($r \leq R_{\text{B}}$), we follow Eqs. (77)–(80) to derive the structural properties of the accretion disk. The default parameters we adopt for the accretion disk correspond to describing a turbulent disk with flux-frozen magnetic fields: $\zeta_{\text{t}} = \zeta_{\text{A}} = 1/3$ and $\psi_{\text{t}}/\psi_{\text{A}} = 1$.

- (xi) Using the properties of the SNF nucleus, we numerically solve the continuity equation, Eq. (84), that describes the evolution of gas profile in the SNF nucleus during the nuclear burst, with specific choices of functional forms detailed in §4.5.2. The default parameters controlling the burst include those describing the AGN feedback to the nucleus, $\xi_{\text{AGN},\infty} = 1$, $v_{\text{out}} = 10^3 \text{ km s}^{-1}$ and $\beta_{\text{out}} = 0$; star formation efficiency in the nucleus, $\alpha_{\text{SF}} = 1$ and $\delta_{\text{Q}} = 1$; and BH accretion efficiency, $\alpha_{\text{BH}} = 1$. The numerical solver, referred to as a ‘refinement engine’ in this paper, works with a time step of the order of a year, much finer than the global integrator so that the $\lesssim \text{Myr}$ timescale of the nuclear burst can be resolved. Examples for the evolution of BH mass, stellar mass, and gas mass during individual nuclear bursts are shown in Fig. 7. Here we take the changes in these masses, $\Delta M_{\text{BH,burst}}$, $\Delta M_{*,\text{burst}}$ and $\Delta M_{\text{g,burst}}$ during the entire burst as outputs of the refinement engine, and add them to the BH mass, stellar mass and gas mass of the bulge of the galaxy, respectively. BH growth through gas accretion in this way is referred to as the ‘wet’ channel in the bursty mode.

- (xii) The energy of AGN feedback released in the nuclear burst could have long-lasting effects on the gas on larger scales, which is also modeled here by a formulation similar to the AGN feedback adopted for the continuous growth of the BH. Specifically, we remove a fraction

$$F_{\text{ej,burst}} = f_{\text{ej}} \left\{ 1 - \exp \left[- \frac{\alpha_{\text{AGN,f}}(z) \Delta M_{\text{BH,burst}} c^2}{M_{\text{b}} V_{\text{g}}^2} \right] \right\} \quad (\text{A15})$$

of gas from that remaining in the bulge, with the parameters $\alpha_{\text{AGN,f}}(z)$ and f_{ej} exactly the same as those used in Eqs. (A7) and (A9), respectively. This implies that the coupling of AGN feedback with the SGC due to a nuclear burst is assumed to be similar to that due to the continuous growth of the BH, which is certainly a simplified treatment and aims to avoid an over-complication of the model.

- (xiii) After a nuclear burst, stars formed in the galactic center may continuously adjust their orbits via dynamical relaxation, causing some stars to migrate to the center and be captured by the BH. The fraction of stellar mass that can sink and be captured depends on the detailed mass and phase-space distribution of the stars, which has not yet been known. To account for this process in a simple way, we move a fraction f_{dry} of stellar mass formed in each nuclear burst ($\Delta M_{*,\text{burst}}$) to the BH. The default value is set to be $f_{\text{dry}} = 6.8 \times 10^{-3}$, derived with analytical approximation by Dekel et al. (2025, see their §3.1.1). We refer to this process of BH growth as a ‘dry’ channel in the bursty mode, as opposed to the ‘wet’ channel modeled in the above step (xi). There should be an order-unity uncertainty in f_{dry} . However, as we have shown in Fig. 14, the dry channel is unimportant compared to the wet channel, and the exact value of f_{dry} thus does not significantly affect our results.
- (xiv) If the subhalo h merges into another subhalo h_{mp} to form a descendant subhalo (i.e. h_{mp} is a main-branch progenitor of the descendant, while h is not), we add a fraction, f_{merge} , of M_{BH} to that of h_{mp} . Similarly, we added a fraction, f_{merge} , of M_* and M_{g} of h to those in the bulge of h_{mp} . The remaining stars and gas are assumed to be tidally stripped or scattered into the halo to become intracluster baryons. Since the merger may also dynamically change the orbits of baryons in the disk of h_{mp} and transform its morphology, we move some stellar and gas masses in the disk of h_{mp} to those in the bulge. The moved masses are assumed to be a fraction, f_{heat} , of M_* and M_{g} of the infalling galaxy hosted by h , since a more massive infalling galaxy is expected to have a larger dynamical effect on the disk. The exact values of f_{merge} and f_{heat} are very uncertain. The reported values in the literature differ significantly (see e.g. Lu et al. 2014; Behroozi et al. 2019), and depend on how the boundaries of galaxies are defined (He et al. 2013; Bernardi et al. 2013). In this paper, we set $f_{\text{merge}} = f_{\text{heat}} = 0.5$ as the default only for demonstrating the effects of mergers. However, even with such a large value, the contribution of mergers to the growth of M_{BH} and M_* is found to be only significant in massive halos at low z (see §5.3.1, Figs. 13 and 14) – a conclusion also reached by hydro simulations (see e.g. figs. 2 and 10 of Li et al. 2024).
- (xv) Gas-phase metallicity of galaxies is modeled according to the ‘gas-regulator’ scenario, with certain modifications accounting for the high- z environment (e.g. cold-mode accretion, high clumpiness) that produces different metal transport from that in the local Universe (see §2.4 of Paper-III). Note that the model of metallicity remains highly uncertain due to large uncertainties in both observations and theories. For example, for the galaxy in the subhalo h , the change of the oxygen mass, ΔM_{Z} , during the current time step is given by the continuity equation that combines the effects of stellar yield, star

formation consumption, and feedback ejection, as

$$\Delta M_{\text{Z}} = y_{\text{eff}} \frac{\Delta M_{*}}{1 - R} - (\Delta M_{*} + \Delta M_{\text{BH}}) Z; \quad (\text{A16})$$

$$y_{\text{eff}} = y f_{\text{mix}} (1 - f_{\text{esc}}); \quad (\text{A17})$$

$$f_{\text{mix}} = \frac{1}{1 + x_{\gamma}}; \quad (\text{A18})$$

$$f_{\text{esc}} = \frac{1 + \alpha_{\text{esc}} x_{\text{v}}}{1 + x_{\text{v}}}. \quad (\text{A19})$$

Here ΔM_{*} and ΔM_{BH} are the change of stellar mass and BH mass, respectively, during the current time step, as modeled in the previous steps; $Z \equiv M_{\text{Z}}/M_{\text{g}}$ is the metallicity of the galaxy; y_{eff} is the effective oxygen yield, which incorporates the intrinsic yield of stars, the escape of decoupled fraction, the ejection by stellar feedback, and the recycling due to cooling, all approximated to be instantaneous; $y = 0.0163$ is the intrinsic oxygen yield of star formation; f_{mix} is the efficiency of metal mixing between the inflow gas and the existing ISM, and $x_{\gamma} = |\gamma/\gamma_0|^{\beta_{\text{mix}}}$ characterizes its dependence on γ_{v} ; f_{esc} is the escaped fraction of metal from the galaxy, $x_{\text{v}} = (V_{\text{max}}/V_{\text{esc}})^{\beta_{\text{esc}}}$ quantifies its dependence on the gravitational potential well h , and $\alpha_{\text{esc}} = z/(z_{\text{esc}} + z)$ characterizes its evolution with redshift. With the above parameterization, the metal mass ejected from the galaxy is

$$\Delta M_{\text{Z,ej}} = f_{\text{esc}} y \frac{\Delta M_{*}}{1 - R}, \quad (\text{A20})$$

Note that the metal mass brought in by inflow gas from the IGM is ignored in Eq. (A16) given that the metallicity in IGM is much lower than that in ISM. The default parameters adopted here are $V_{\text{esc}} = 75 \text{ km s}^{-1}$, $\beta_{\text{esc}} = 2$, $z_{\text{esc}} = 5$, $\gamma_0 = 1$, and $\beta_{\text{mix}} = 2$.

With the above steps, we can update the global properties – such as the masses of the BH, stars and gas – of the galaxy in h during the current time step. The growth of these mass components contributed by different channels can be separated, as shown in Figs. 13 and 14 for examples of halos. The evolution of galaxies also provides the basis for modeling their radiation and matter feedback to the large-scale environment, as to be detailed in the following.

A5 The model of Lyman-Werner radiation field

As shown in §5.2.3 and Fig. 11, the radiation feedback in the LW band from galaxies to their large-scale environment is a critical factor that determines when and how the first generations of stars and BHs in halos illuminated by the radiation can form. An accurate modeling of the radiation field requires full spectral information from galaxies and precise algorithm for radiative transfer, both of which are computationally prohibitive in a cosmological volume. Here we describe a simplified prescription suitable for our purpose.

We first note that the modeling of radiation feedback can take advantage of the characteristics of the LW photons: these photons have a large mean free path, up to $\approx 100 h^{-1} \text{ Mpc}$ at $z = 20$ (see Eq. A24 below and Ahn et al. 2009), comparable to the spatial scale of the box of the N-body simulation used in this paper. The intensity of LW radiation can exhibit strong variation (both spatially and temporally), due to the clustering (i.e. biased distribution) of galaxies in the cosmic volume, the strong temporary variation of star formation rate of high- z galaxies (especially when our bursty mode is included; see e.g. Fig. 14), and the sharp (r^{-2} ; see below) radial decay from individual sources.

To account for both the large mean free path and the strong variation in spacetime of LW radiation, we follow Agarwal et al. (2012) to decompose $J_{\text{LW}}(\mathbf{x}, t)$, the total LW intensity illuminating at an

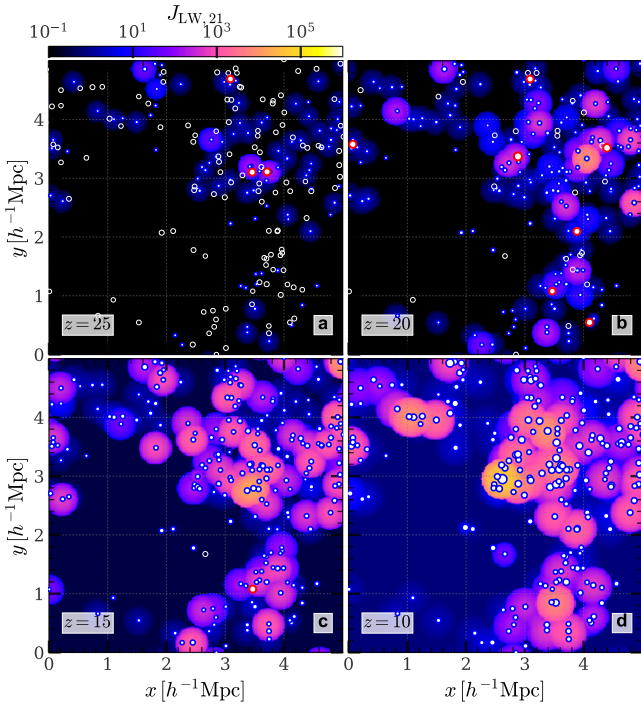


Figure A2. Lyman-Werner (LW) radiation field predicted by the model. **a–d**, LW intensity, $J_{\text{LW}} = J_{\text{bg}} + J_{\text{local}}$, color coded according to the color bar on the top left, in a $5 \times 5 \times 2 (h^{-1} \text{cMpc})^3$ slice of the simulation box at four different redshifts. The slice is a part of that shown in Fig. 12g–i (lower-left corner in x - y , and middle slice along z). Each dot represents a subhalo, sized proportionally to R_v , and styled according to the status of the galaxy it hosts: seeded (blue edge), unseeded and unilluminated ($J_{\text{LW},21} < 1$, white edge), unseeded and illuminated ($J_{\text{LW},21} \geq 1$, red edge). LW photons represent a type of radiation feedback from galaxy formation to the large-scale environment, and play a critical role in delaying the formation of the first generations of stars and BH seeds in subhalos illuminated by the radiation. See Appendix A5 for the detailed steps to model the LW radiation.

arbitrary location \mathbf{x} and a cosmic time t , into a global term of radiation background, J_{bg} , and a local term, J_{local} , mainly contributed by nearby star-forming sources, as:

$$J_{\text{LW}}(\mathbf{x}, t) = J_{\text{bg}}(t) + J_{\text{local}}(\mathbf{x}, t). \quad (\text{A21})$$

The background term includes the contribution from all radiation sources in the universe, and can be obtained by using the total stellar mass density $\rho_*(t)$ of all survived stars in the simulated volume, as

$$J_{\text{bg}}(t) = f_{\text{esc}} \frac{hpc}{4\pi m_p} \eta_{\text{LW}} \rho_*(t) (1+z)^3. \quad (\text{A22})$$

Here, $f_{\text{esc}} \leq 1$ is the fraction of LW photons that escape from galaxies; h_p is the Planck constant; c is the speed of light; m_p is the proton mass; η_{LW} is the number of LW photons emitted per proton mass; $z = z(t)$ is the redshift at time t ; the factor $(1+z)^3$ accounts for the expansion of the universe. The background is uniform and isotropic, as implied by the cosmological principle. This formulation implies that ρ_* obtained in the simulated volume is a fair representation of the entire universe, which is satisfied as long as we are using a simulation with a large volume and a fairly sampled initial condition (§2). The default parameters we use are $f_{\text{esc}} = 1$ (Dijkstra et al. 2014; Spinoso et al. 2023), assuming that the LW emitters have been depleted of H_2 by recent star formation, and $\eta_{\text{LW}} = 4 \times 10^3$ (Agarwal et al. 2012) for Pop-II stars. We have not incorporated the contribution from Pop-III stars, since their star formation efficiency (SFE) is low (see §3.2), and

the post-seeding growth in the bursty mode dominates the growth of galaxies soon after the seeding (see §5.3.1 and Fig. 14). We have also not incorporated the contribution from AGNs, but we note that the recently observed high number density of faint AGNs at high z suggests that they should be seriously taken into account (see e.g. Fig. 5.3.2 for the cosmic BHAR density).

The local term is modeled by summing up the contribution from individual source galaxies with a spherically symmetric approximation for the emission, as

$$J_{\text{local}}(\mathbf{x}, t) = \sum_i \frac{f_{\text{esc}}}{4\pi \Delta\nu_{\text{LW}}} \frac{Q_{\text{LW}} M_{*,\text{alive},i}}{4\pi d_{\text{L},i}^2} f_{\text{mod},i}. \quad (\text{A23})$$

Here, Q_{LW} is the LW luminosity per unit alive stellar mass; $\Delta\nu_{\text{LW}}$ is the frequency range of LW radiation (11.2–13.6 eV); $M_{*,\text{alive},i}$ is the mass of alive stars in the i -th source galaxy; $d_{\text{L},i}$ is the luminosity distance between the source galaxy and the target location \mathbf{x} ; $f_{\text{mod},i}$ is a modulation factor that accounts for the attenuation of LW radiation by IGM. To avoid over-complication of radiative transfer, we use the fitting formula obtained by Ahn et al. (2009) for $f_{\text{mod},i}$. Adapting it to our cosmology, we obtain

$$f_{\text{mod},i} = 1.7 \exp \left[- \left(\frac{d_{\text{c},i}}{110 \alpha_{\text{mod},i} h^{-1} \text{Mpc}} \right)^{0.68} \right] - 0.7 \quad (\text{A24})$$

for $d_{\text{c},i} \leq 92.4 \alpha_{\text{mod},i} h^{-1} \text{Mpc}$, the LW horizon, and $f_{\text{mod},i} = 0$ otherwise. Here, $d_{\text{c},i}$ is the comoving distance between the source galaxy and \mathbf{x} , the scaling factor $\alpha_{\text{mod},i}$ is

$$\alpha_{\text{mod},i} = \left(\frac{\Omega_{\text{m},0}}{0.3089} \right)^{-1/2} \left(\frac{1+z_i}{10} \right)^{-1/2}, \quad (\text{A25})$$

and z_i is the redshift of the source galaxy. The default value we adopt is $Q_{\text{LW}} = 10^{36} \text{ erg} \cdot \text{s}^{-1} M_{\odot}^{-1}$, obtained by averaging the emission of massive O/B stars over their lifetime of $\approx 10 \text{ Myr}$ (Dijkstra et al. 2014, see their Eq. 6). The corresponding mass, $M_{*,\text{alive},s}$, is computed by including only these young massive stars. The summation should be performed over all galaxies in the past light cone of (\mathbf{x}, t) , leading to a computational complexity of $\mathcal{O}(N^2)$ that is prohibitive for a large number (N) of galaxies. To proceed, we first note that the light speed is large enough so that the emission and reception of LW radiation between two subhalos in the simulated volume can be simplified as synchronous events. Thus, the summation can be performed by using sources within the current snapshot. This also leads to an approximation $d_{\text{L},i} \approx d_{\text{c},i}$. To reduce the number of summations, we limit the zone of influence of each source galaxy to a sphere with a radius at which the LW intensity emitted by this galaxy drops to a negligible threshold, $J_{\text{local,th}}$. The default value we adopt is $J_{\text{local,th},21} = 0.01$, which is about a factor of 10^{-2} – 10^{-3} smaller than $J_{\text{crit},21}$, the value to produce a significant delay in gas collapse (see §3.1.3; see also Appendix A of Spinoso et al. 2023 for a similar approach). With these simplifications, the computational time for the LW intensity in a single time step is reduced to a few seconds for our sample of subhalos (§2).

Fig. A2 shows the LW intensity, J_{LW} , in a slice of the simulation box at four different redshifts, $z = 25, 20, 15$ and 10 . Subhalos with seeds having been bred are represented by dots with blue edges, which are the sources of LW radiation. Subhalos without seeds are represented by dots with a white ($J_{\text{LW},21} < 1$) or a red edge ($J_{\text{LW},21} \geq 1$). At $z \approx 25$, the universe is in the seeding era (see §5.3.2), and most subhalos are still on their way to form the first generation of stars and BH seeds. The large number of unseeded subhalos in panel a reflects this infancy era of galaxy formation. The seeding era ends

at $z \approx 20$ when most subhalos have been seeded, which can be seen from the reduction in the number of unseeded subhalos from panel a to panel b. The rapid post-seeding star formation through the bursty mode produces intense but localized LW radiation from the seeded galaxies, manifested as individual lamp-like illuminated zones in panel b. A small fraction of unseeded subhalos located in these zones are illuminated by the LW radiation, and are delayed in forming their first generation of stars and BH seeds. At $z \lesssim 15$, the rising of cosmic star formation rate density (see also Fig. 5.3.2) lifts the global background. However, by this time, most subhalos have already been seeded (see also Fig. 5.2.2), providing harbors of star formation and BH growth that are well shielded from external radiation.

A6 The model of intergalactic-medium enrichment

The enrichment of the IGM is a type of matter feedback of galaxy formation to the large-scale environment. Compared with the radiation feedback described above, matter feedback also plays a role in determining the formation of seeds in subhalos surrounded by the enriched IGM; a major difference is that the diffusion of metal is much slower than that of radiation, so that the expansion of the zone of enrichment produced by the matter feedback from any galaxy has to be tracked over a long timescale.

To simplify the model of IGM enrichment so that it can be efficiently implemented into a large volume, we make the following assumptions: the intergalactic diffusion of metal is assumed to be a spherically symmetric process around each source galaxy, and the enriched volume is described by a spherical bubble; the energy powering the expansion of the bubble is assumed to come from the blast wave generated by SNe, while the contribution from Pop-III stars and AGNs is ignored. All these assumptions are consistent with, and explained in our treatment for the radiation feedback (Appendix A5).

To estimate the sizes of SN bubbles produced by galaxies, we adopt a self-similar solution of blast wave (see e.g. §8.6 of Mo et al. 2010):

$$R_{\text{sh}} = A_{\text{sh}} \left(\frac{E_{\text{SN}}}{\rho_{\text{g}}} \right)^{1/5} (t - t_0)^{2/5}, \quad (\text{A26})$$

where R_{sh} is the physical radius of the shock wave, $A_{\text{sh}} \approx 1$ is a normalization constant whose value is determined by energy conservation, E_{SN} is the energy released by SNe during a burst of star formation, ρ_{g} is the density of the IGM swept by the shock wave, and $t - t_0$ is the elapsed time since bubble generation at time t_0 . We express the energy released by SNe as $E_{\text{SN}} = \epsilon_{\text{SN}} R \Delta M_{*,\text{new}}$, where the energy efficiency $\epsilon_{\text{SN}} = 10^{51} \text{ erg} \cdot \text{M}_{\odot}^{-1}$ (Dijkstra et al. 2014; Spinoso et al. 2023), R is the fraction of returned mass (Eq. A1), and $\Delta M_{*,\text{new}}$ counts all stars newly formed in the past duration of t_{V} (Eq. 7) in which the replenishment of the gas may be accomplished to trigger the next episode of star formation. If there are multiple bubbles generated in the main branch of a galaxy, we only consider the largest (Ventura et al. 2024). However, since SFRs of high- z galaxies often increase rapidly with time (see examples in Fig. 14), the largest bubble is most likely produced in the recent star-forming episode.

Before we implement the IGM enrichment model in our pipeline, we provide an analytical estimate for the bubble sizes. We take $\rho_{\text{g}} = \epsilon_{\rho} \rho_{\text{b}}(z)$, where ρ_{b} is the cosmic baryon density (Eq. 1), and $\epsilon_{\rho} = 60$ (Dijkstra et al. 2014), roughly the density at R_{V} of any halo at z . The

stellar mass formed in the recent t_{V} is:

$$\begin{aligned} \Delta M_{*,\text{new}} &= \epsilon_{\text{SF}} f_{\text{b}} \dot{M}_{\text{V}} t_{\text{V}} \\ &= \left[1.73 \times 10^7 \text{ M}_{\odot} \right] \epsilon_{\text{SF},0.1} M_{\text{V},9.5}^{1.14} (1+z)_{10}, \end{aligned} \quad (\text{A27})$$

where $\epsilon_{\text{SF}} = 0.1 \epsilon_{\text{SF},0.1}$ is the star formation efficiency, with its typical value of 0.1 taken from observations (e.g. Wang et al. 2025) and empirical models for high- z galaxies (e.g. Nusser 2024), and \dot{M}_{V} is the halo accretion rate, approximated using Eq. (8). Substituting the above equation into Eq. (A26) and taking $t - t_0 = t_{\text{V}}$, we can obtain the bubble radius in comoving units as

$$R_{\text{sh},\text{c}} = (1+z) R_{\text{sh}} = 231 \text{ kpc} \left[\chi_{\text{sh}} M_{\text{V},9.5}^{1.14} \right]^{1/5}, \quad (\text{A28})$$

where we have defined $\chi_{\text{sh}} = \epsilon_{\text{SN},51} \epsilon_{\text{SF},0.1} R_{0.4} / \epsilon_{\rho,60}$, $\epsilon_{\text{SN},51} = \epsilon_{\text{SN}} / (10^{51} \text{ erg} \cdot \text{M}_{\odot}^{-1})$, $R_{0.4} = R/0.4$ and $\epsilon_{\rho,60} = \epsilon_{\rho}/60$. Thus, the comoving size of an SN bubble is only a weak function of halo mass, and is redshift insensitive, if χ_{sh} does not vary significantly. Another useful quantity is the ratio between R_{sh} and R_{V} (Eq. 3) of the host halo,

$$R_{\text{sh}}/R_{\text{V}} = 5.05 \chi_{\text{sh}}^{1/5} M_{\text{V},9.5}^{-0.105}, \quad (\text{A29})$$

which again depends weakly on the halo mass and is largely insensitive to redshift, suggesting that the radius of an SN bubble is nearly proportional to the size of its host halo.

We can also estimate the global volume filling factor, f_{V} , of SN bubbles (i.e. the fraction of cosmic volume filled by the bubbles). For this, we multiply the volume of bubble (in comoving unit) with $\Phi(M_{\text{V}})$, the halo mass function (defined similarly to the mass function of BH seed in Eq. 101), and integrate it over the full range of halo mass:

$$f_{\text{V}} = \int \frac{4}{3} \pi R_{\text{sh},\text{c}}^3 \Phi(M_{\text{V}}) \text{d} \log M_{\text{V}}. \quad (\text{A30})$$

Since the power index on M_{V} in the halo mass function is smaller than -0.8 (see §3.1 of Tinker et al. 2008), the integrand scales with halo mass as $R_{\text{sh},\text{c}}^3 \Phi \sim M_{\text{V}}^{\beta}$, with $\beta < 1.14 \times 3/5 - 0.8 \approx -0.12 < 0$. Thus, the volume filling factor is dominated by low-mass halos. This signifies the need for modeling the full population of galaxies in a cosmological volume, particularly in the early universe where halo masses are typically small, so that the environment that affects the seeding can be fully captured.

In our numerical implementation, we follow Ventura et al. (2024) to model the spatial distribution of metal on coarse grids, and ignore the sub-grid distribution. The reason for adopting such an approach is to take into account the limitations arising from the simplifications we have made: (i) the assumptions of spherical symmetry and blast-wave expansion are not able to catch the sub-structure of bubbles; (ii) interactions among bubbles generated by different galaxies or by one galaxy at different episodes of star formation cannot be resolved; (iii) the locations of subhalos extended by our Monte Carlo algorithm (Appendix A1) are not exactly known. Thus, a grid-based approach with spatial resolution comparable to these physical limitations is sufficient for our purpose. The steps of our numerical implementation are as follows:

- (i) We partition the simulation box into cubic cells with a fixed side length (in comoving unit) that is sufficiently fine compared with the spatial scale to be used to smooth the spatial distributions (see step iv below). Any field in the simulation box is then defined by a set of discrete values in the cells. The first field to compute is ρ_{m} , the mass density field of the current snapshot, whose value in each cell C is obtained by summing up the masses of all the simulated dark matter particles located in C , and dividing it by ΔV_{C} , the volume of the cell.

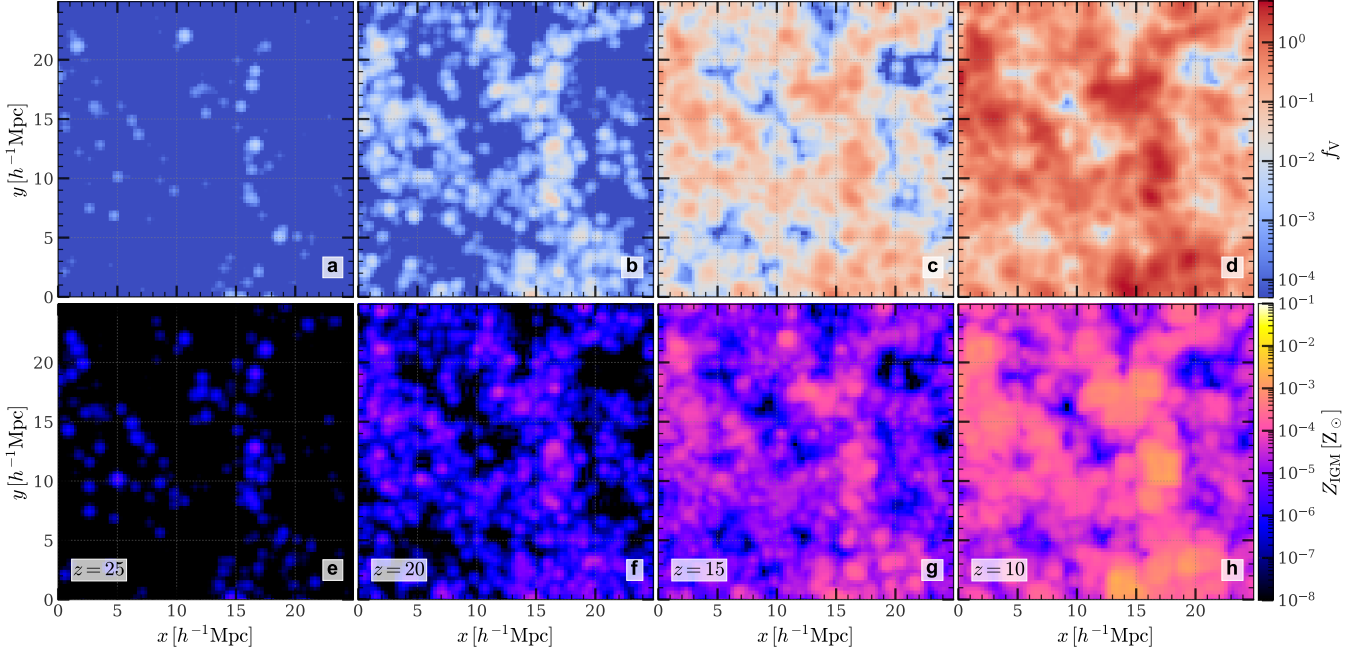


Figure A3. Enrichment of intergalactic medium (IGM) predicted by the model. **a–d**, distribution of volume filling factor (f_V) of SN bubbles in a $25 \times 25 \times 2 (h^{-1}\text{cMpc})^2$ slice of the simulation box at four different redshifts. **e–h**, similar but for Z_{IGM} , the metallicity of IGM. Note that Z_{IGM} shown here is the average value within each entire cell, not just within the bubble-filled fraction (i.e. $Z_{\text{IGM}} = \Delta M_Z / (f_b \rho_m \Delta V_C)$, without an f_V in the denominator). For both f_V and Z_{IGM} , mean values along z -axis are taken, and are color coded according to the color bar on the right. The slice shown here is exactly the middle $2 h^{-1}\text{Mpc}$ along z -axis of that shown in Fig. 12g–i. At $z \approx 20$, a noticeable fraction of the cosmic volume has been enriched by metals ejected from early-formed galaxies, leading to the significant formation activity of Pop-II seeds; by the time of $z \approx 10$, most cosmic volume has already been enriched to the Pop-II level, marking a transition of seeding channel from Pop-III to Pop-II (see e.g. Fig. 10 for the cosmic seeding history). See Appendix A6 for the detailed steps to model the IGM enrichment.

- (ii) For each subhalo, we track the evolution of R_{sh} of the SN bubble generated by the galaxy hosted by the subhalo, and update it over time steps. In the current snapshot s , we compute $\Delta M_{*,\text{new}}$, the mass of newly formed stars over the past duration of t_V in the subhalo $h \in H_S$. We compare $\Delta M_{*,\text{new}}$ with those evaluated in the previous time steps along the main branch of h , and take the largest value among them to power the expansion of the SN bubble. We find the derivative of R_{sh} with respect to t using Eq. (A26), as

$$\dot{R}_{\text{sh}} = \frac{2}{5} A_{\text{sh}} \left(\frac{E_{\text{SN}}}{\rho_g} \right)^{1/5} (t - t_0)^{-3/5}. \quad (\text{A31})$$

Here E_{SN} is computed using the largest value of $\Delta M_{*,\text{new}}$ and t_0 is the time when this value is achieved. The gas density ρ_g is evaluated at halo-centric distance R_{sh} by assuming an isothermal profile, and a total mass of $f_b M_V - M_g - M_* - M_{\text{BH}}$ (i.e. the available baryon mass subtracted by the gas, stellar and BH masses in the galaxy hosted by h) for the gas within h (§4.1 and Fig. 6), if R_{sh} is smaller than R_V of h ; otherwise, we use ρ_m at the location of h on a scale of R_{sh} to find the gas density as $\rho_g = f_b \rho_m$. We then use \dot{R}_{sh} to update R_{sh} for one time step.

- (iii) We define the volume filling factor f_V of any cell as the fraction of the cell volume that has been filled by SN bubbles, and we define the metal mass ΔM_Z in any cell as the total mass of metal ejected by galaxies into the cell. To find these fields for a subhalo $h \in H_S$, we identify the intersection of its SN bubble with any cell, and accumulate the contribution from h to f_V of the cell by using the volume of the intersection; similarly, we compute the total mass of metal ejected from h and all of its progenitors, assuming this mass is

uniformly distributed within the SN bubble of h , and accumulate the contribution from h to ΔM_Z in any cell intersected with the bubble.

- (iv) We smooth the fields, ρ_m , f_V and ΔM_Z by a Gaussian kernel with a comoving size of $l_{\text{sm}} = 250 h^{-1}\text{kpc}$, a scale comparable to the typical size of SN bubbles estimated in Eq. (A28). We compute the metallicity of the IGM in the bubble-filled volume of any cell as $Z_{\text{IGM}} = \Delta M_Z / (f_b f_V \rho_m \Delta V_C)$, and as $Z_{\text{IGM}} = 0$ outside the bubble-filled volume.
- (v) For any subhalo $h \in H_S$, the metallicity of the IGM around the galaxy hosted by h is determined by a stochastic algorithm. Specifically, we draw $u_Z \sim U[0, 1]$, a random number with a uniform distribution over $[0, 1]$, and compare it with the value of f_V in the cell containing h . If $u_Z \leq f_V$, the galaxy is considered to be located in the bubble-filled volume of the cell and the value of Z_{IGM} is assigned accordingly; otherwise, it is located outside the bubble-filled volume and $Z_{\text{IGM}} = 0$ is assigned.

Fig. A3 shows the volume filling factor and IGM metallicity of the cells in a slice with a thickness of $2 h^{-1}\text{Mpc}$ (in comoving units) in the simulation box at four different redshifts, $z = 25, 20, 15$, and 10 . At $z = 25$ (panels a and e), a few SN bubbles can be seen around galaxies whose post-seeding star formation has just started owing to the onset of nuclear bursts (see Fig. 14 for example galaxies). The sizes of these bubbles are small, as expected from the low halo masses (see e.g. Fig. 11) and the proportional relation between R_{sh} and R_V (Eq. A29). The metallicity of the enriched volume is as low as $\sim 10^{-6} Z_{\odot}$, which is a consequence of the low star formation efficiency in low-mass halos (see e.g. Fig. 15b for a comparison between $\dot{\rho}_*$ and virialized $\dot{\rho}_b$). From $z = 25$ to $z = 20$ (panels b and f), the bubble-filled volume expands significantly from the sites densely populated with galaxies

that have been newly seeded and experiencing intense star formation through the bursty mode. The metallicity in some bubble-filled zones increases beyond $10^{-4} Z_{\odot}$, and the unseeded subhalos swept by the bubbles have increased chances to be fueled with enriched gas and to form Pop-II stars and BH seeds. The rapid increase of cosmic seeding activity contributed by Pop-II seeds (Fig. 10a, green shade) at $z \gtrsim 20$ can thus be explained. At $z \lesssim 20$, most subhalos had already been seeded, while their SFRs through the bursty mode continue to increase due to the fast halo assembly (Fig. 1a). The SN bubbles continue to expand and overlap with each other to form extended structures, leaving only a small fraction of the cosmic volume pristine (panels c and g). By the time of $z = 10$ (panels d and h), most cosmic volume has been filled by SN bubbles, with some hot spots having $Z_{\text{IGM}} \gtrsim 10^{-3} Z_{\odot}$. In this era, most of the newly bred seeds are Pop-II, as can be seen by a comparison between the cosmic seeding rate densities of all seeds and of Pop-II seeds alone (Fig. 10a).

APPENDIX B: EVOLUTION OF THE STRUCTURE HIERARCHY

In §4.1, we have described the structure hierarchy (Fig. 5) that leads to BH growth and shown the ladder-like gas profile for a halo of MW size and its progenitors at different redshifts (Fig. 6). Here we present the properties of individual levels on the ladder, and the evolution of these properties, in the histories of halos. In application, one can start from any level on the ladder, use it as a large-scale boundary condition for the processes that lift gas to higher levels, and resolve the formation and growth of structures on smaller scales.

Fig. B1 shows the evolution of the properties of SGC and SNF nucleus in the assembly histories of halos with different masses at $z = 0$ ($M_{v,z=0}$). To produce these properties, we first take the median assembly history of halos with a given $M_{v,z=0}$ in the same way as that used to obtain Fig. 1. We then assume that a global disturbance has been induced in the SGC hosted by a halo at a redshift z , and that the disturbance generates an SNF nucleus in the center of the SGC. We compute the properties of the SGC according to Eqs. (9)–(13) (the first equality in each of them) and the properties of the SNF nucleus according to Eqs. (40)–(49). We perform the calculation for halos with any $M_{v,z=0}$ and at any z , and show the gas mass (M_g), radius (R_g), surface density (Σ_g) and circular velocity (V_c) in panels a1–d1, respectively, for both SGC (solid) and SNF nucleus (dashed). For reference, we also show the ratio between the properties of SNF nucleus and SGC in panels a2–d2. The gas fraction assumed here is $f_{g,0.157} = 1$, which describes a gas-rich SGC that has not been depleted by feedback; the radius fraction is $f_{r,0.04} = 1$, roughly the critical value at which rotation is still unable to provide enough support; the feedback-free threshold is $n_{\text{snf},3.5} = 1$, which corresponds to a free-fall timescale of about 1 Myr. Halos with values of $f_{g,0.157}$, $f_{r,0.04}$ or $n_{\text{snf},3.5}$ other than 1 can also be analyzed by the equations listed above. Note that at $z \gg 1$ (when the universe is approximately an Einstein-de Sitter universe), the dependence of SGC properties on M_v and z can be analytically approximated using Eqs. (9)–(13); if, in addition, $z \lesssim 10$ (i.e. the concentration $c_g \equiv R_{\text{SGC}}/R_{\text{nuc}} \gg 1$), the dependence of the properties of SNF nucleus on M_v and z can also be analytically approximated using Eqs. (51)–(56).

The most obvious feature of the SNF nucleus compared to the host SGC is the redshift evolution. At $z \gtrsim 20$ (when the age of the universe is less than 200 Myr), an SGC is entirely supernova-free, as seen from the ratio of M_g (or R_g) between the SNF nucleus and the SGC. At lower redshift, the SNF nucleus occupies less fraction

of the SGC: at $z \approx 5$, the SNF nucleus takes about 10% of the SGC mass, and at $z = 0$, the fraction drops to about 1% (panel a2); a similar trend is seen for the radius (panel b2). The surface density of the SNF nucleus, however, remains higher than $10^3 M_{\odot} \text{pc}^{-2}$ at any redshift for halos with any mass shown in Fig. B1, and the ratio of Σ_g between SNF nucleus and SGC continuously increases with time. The evolution of the properties of SNF nucleus and SGC reflects that the importance of SNF nucleus in supplying gas to the central BH of a galaxy declines with time. This, together with the less frequent global disturbance at lower redshift, explains the decrease of BH growth through the bursty mode with time (see §5.3.1 and Fig. 14).

Climbing one level higher on the ladder, one can deduce the Bondi properties of the central BH hosted by an SNF nucleus. The properties most relevant to the modeling of BH growth are the Bondi radius (R_B ; Eq. 60) and the Bondi accretion rate (\dot{M}_B or $\dot{m}_B \equiv \dot{M}_B/\dot{M}_{\text{Edd}}$; Eq. 59), which forms the boundary condition to determine the structure of the accretion disk around the BH that experiences the bursty mode accretion. In Fig. B2, we show the evolution of R_B and \dot{M}_B in panels a and b, respectively, for a BH with a mass of $M_{\text{BH}} = 10^4 M_{\odot}$ (dashed) or $10^6 M_{\odot}$ (solid) placed in the center of an SNF nucleus in the history of a halo with a given $M_{v,z=0}$. At $1 \ll z \lesssim 10$, R_B and \dot{m}_{BH} can also be approximated analytically by linking them to the properties of either the SNF nucleus, or the SGC, or the halo, using Eqs. (68) and (69), respectively. As expected from these approximations, R_B increases with M_{BH} , reflecting the larger zone of influence of more massive BHs, and decreases with M_v , due to the deeper gravitational potential well in the centers of more massive halos. On the other hand, \dot{M}_B increases with both M_{BH} and M_v . Note that in some cases where a too massive BH is placed in a too small SNF nucleus, R_B is obtained by extrapolating the SGC density profile (Eq. 44) outward beyond R_{nuc} , which can lead to $R_B/R_{\text{nuc}} > 1$. In realistic applications where BHs grow from well-defined initial and boundary conditions, such cases should not occur.

APPENDIX C: VARIANTS OF THE MODEL

As described §5.3.1, the growth of stellar and BH masses in galaxies are modeled as a multichannel process, with each channel governing a certain stage of the growth. To cover the broad scales of time, space and mass involved in the growth, we have incorporated certain assumptions for each channel, and used the parameters listed in Table A1 to characterize the degree of freedom (DoF) in the model. Some parameters have already been assigned values or ranges based on either analytical estimates, numerical simulations or observational/empirical constraints, while others remain uncertain. Here, we perform a set of controlled experiments by introducing a number of variants of the model. The purpose is neither to perform an exhaustive exploration in the full parameter space nor to find a best-fit calibration, but to demonstrate the effects of different physical recipes on model predictions for the stellar and BH contents of galaxies and to explore implications for future observations.

To maximize the coverage of the parameter space, the variants to be considered include adaptations for each of the growth channels. To minimize the entanglement of the effects of different processes, we limit ourselves to changing the parameters of only one process at a time to produce a new variant. Note that the modifications made on one process may affect the consequences of other processes. For example, a change in the star formation efficiency in any channel also modifies the LW radiation field and IGM enrichment, thus affecting the seeding in other halos affected by the radiation or matter feedback. To avoid such complication, new variants introduced here are

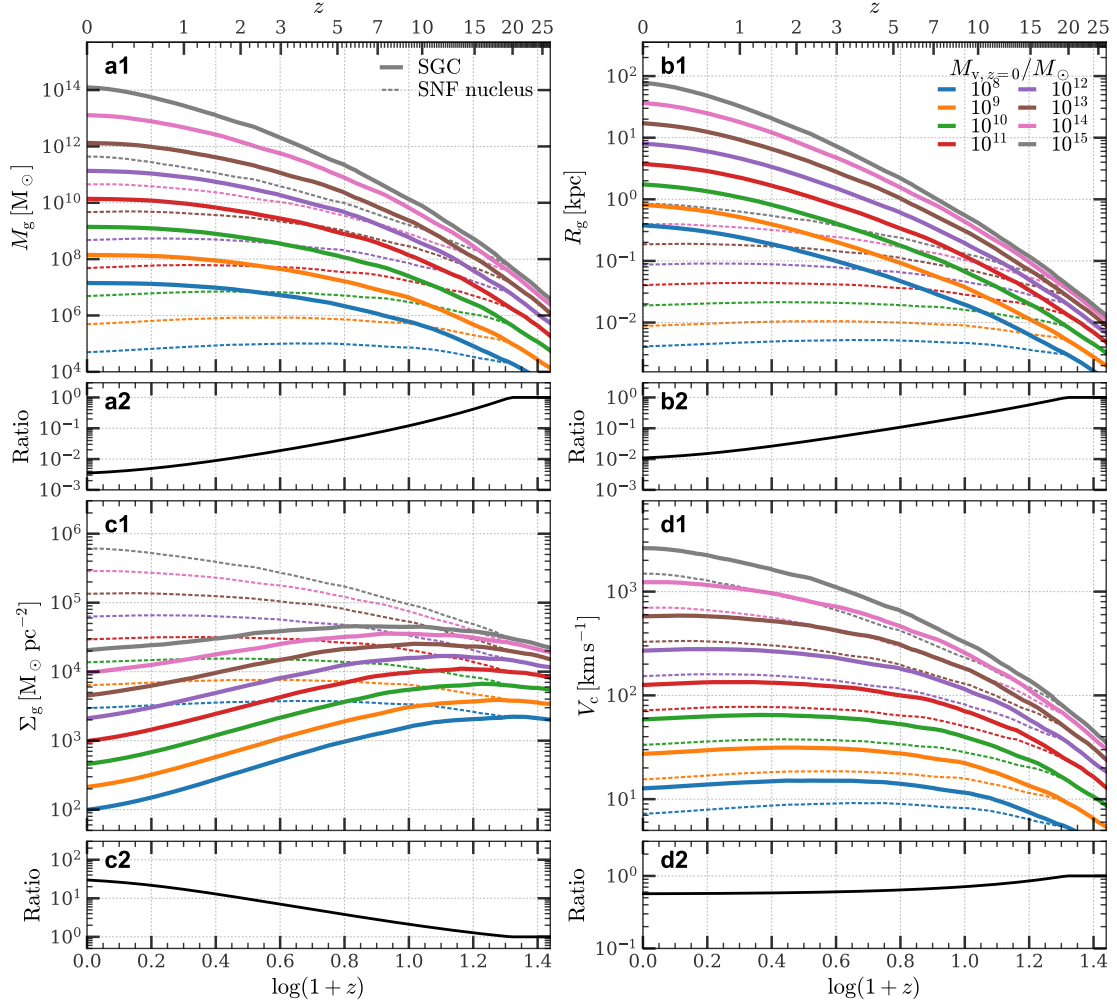


Figure B1. Properties of SGC and SNF nucleus in the histories of halos with different masses. Four properties for SGC (solid) and SNF nucleus (dashed) are shown: gas mass (M_g ; **a1**), gas radius (R_g ; **b1**), gas surface density (Σ_g ; **c1**), circular velocity (V_c ; **d1**). **a2–d2** show the ratios of the corresponding properties between SNF nucleus and SGC. Each property is shown as a function of redshift evaluated based on median assembly history of halos with a given mass M_v at $z = 0$. The assembly histories of halos used in this figure are exactly the same as those shown in Fig. 1. Analytical approximations for the SGC properties can be found in Eqs. (9)–(13), and for the SNF nucleus properties can be found in Eqs. (51)–(56). See Appendix B for details.

based on the ‘MinSeed’ variant (see §5.3.3), which pins down the seeding process by always producing minimal masses of BH seeds in mini-halos. All variants are applied to the same set of merger trees (obtained from TNG100-1-Dark; see §2) as used for the main texts.

The variants considered here, including those defined in the main texts, are listed below.

- (i) ‘*Default*’: this is the default case used to produce the results in the main texts (except those shown in Fig. 16), with the values of parameters listed in Table A1.
- (ii) ‘*MinSeed*’: this variant is the same as ‘Default’, except that Pop-III seeds with minimal masses are forced to form in all mini-halos. To achieve this, we set the parameter $\gamma_{\text{crit}} \rightarrow +\infty$ and force $J_{\text{LW}} = 0$ and $Z_{\text{IGM}} = 0$ for all halos (see §5.3.3).
- (iii) ‘*MaxSeed*’: this variant is the same as ‘Default’, except that Pop-III seeds with maximal masses are forced to form in all mini-halos. This is achieved by setting the parameter $\gamma_{\text{crit}} \rightarrow 0$, and force $J_{\text{LW}} = +\infty$ and $Z_{\text{IGM}} = 0$ for all halos (see §5.3.3).
- (iv) ‘*NoBurst*’: this variant is the same as ‘MinSeed’, except that nuclear burst is completely turned off. This is achieved by setting $\delta_\gamma = +\infty$ (see Eq. 42), so that the formation of SNF nucleus never occurs.

- (v) ‘*BurstHighSFE*’: this variant is the same as ‘MinSeed’, except that the star formation efficiency in nuclear burst is boosted, so that the competition between star formation and BH accretion is tilted towards the former and thus BH cannot grow by the wet channel in the bursty mode. This is achieved by setting $\alpha_{\text{BH}} = 0$ and $\delta_Q \rightarrow 0$ (see §4.5.2). Note that the dry channel in the bursty mode is still open in this variant.
- (vi) ‘*AGNConstEff*’: this variant is the same as ‘MinSeed’, except that the efficiency of AGN feedback is set to be redshift-independent. This is achieved by setting $\beta_{\text{AGN},f} = 0$ (see Eq. A7).

Fig. C1 shows M_{BH} and M_* predicted by the above model variants at different redshifts. To limit the dynamic range of displayed values, we show the ratio, M_{BH}/M_* , instead of M_{BH} itself, as a function of M_* . Obvious differences can be seen among the variants with adaptations for the seeding methods (Default, MinSeed and MaxSeed). At $z = 20$ (panel a), about the ending point of the seeding era (see §5.3.2 and Fig. 15), the MaxSeed (MinSeed) variant predicts the highest (lowest) M_{BH}/M_* at any M_* , and the Default variant predicts a value in between the two extremes. The largest value of M_* produced by the MaxSeed is also the largest among the three variants, while that

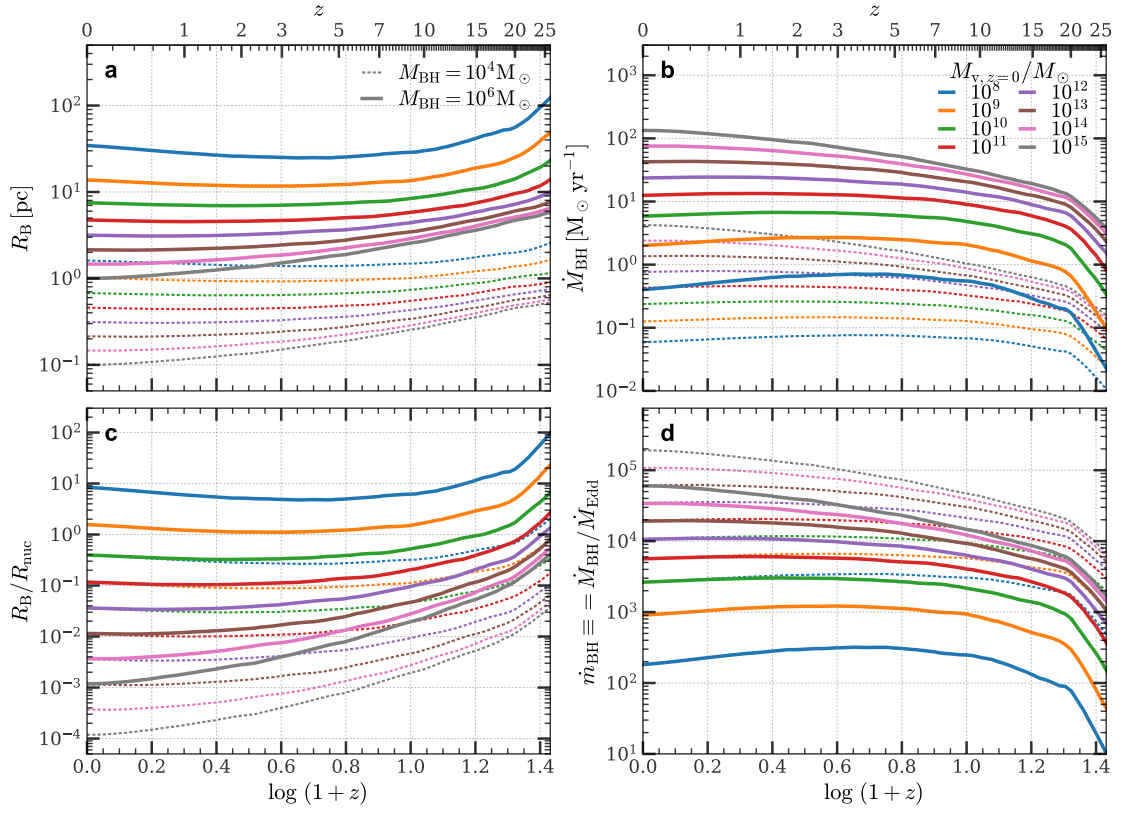


Figure B2. Bondi properties of central BHs in the histories of halos with different masses. Here we show four properties: Bondi radius (R_B ; **a**), Bondi accretion rate (\dot{M}_B ; **b**), the ratio between R_B and the radius of the SNF nucleus hosting the BH (**c**), and the Eddington ratio of Bondi accretion rate ($\dot{m}_{BH} \equiv \dot{M}_B/\dot{M}_{Edd}$; **d**). Each property is shown as a function of redshift evaluated based on properties of the host SGC and SNF nucleus (Fig. B1) in halos with a given mass M_v at $z = 0$, assuming a BH with a mass of $M_{BH} = 10^4 M_\odot$ (**dashed**) or $10^6 M_\odot$ (**solid**) is placed at galactic center. See Appendix B for details.

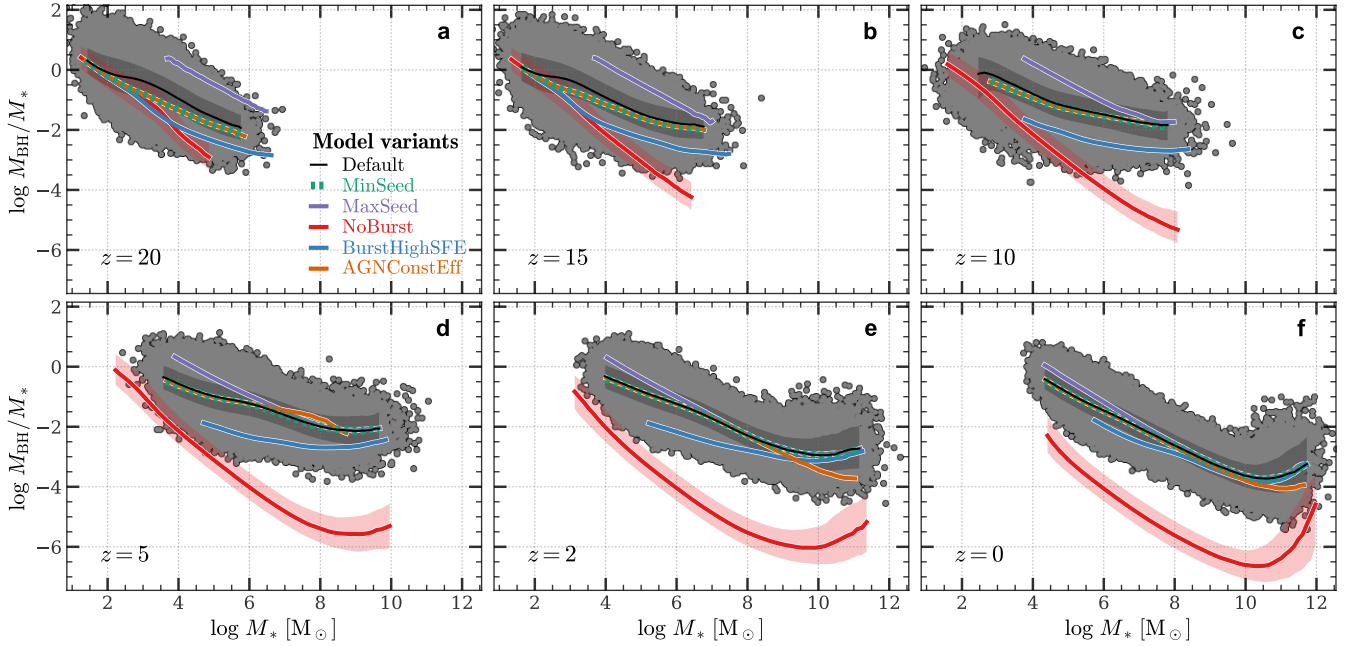


Figure C1. M_{BH}/M_* - M_* relations predicted by the variants of the model. **a-f**, M_{BH}/M_* versus M_* predicted at different redshifts. In each panel, curves show the median values of M_{BH}/M_* as a function of M_* predicted by six variants (**black**, ‘Default’; **green dashed**, ‘MinSeed’; **purple**, ‘MaxSeed’; **red**, ‘NoBurst’; **blue**, ‘BurstHighSFE’; **orange**, ‘AGNConstEff’). **Scatter** points show the predictions for individual galaxies by the ‘Default’ variant. For ‘Default’ and ‘NoBurst’, we also show the $1-\sigma$ range (16th–84th percentiles) by shaded regions. The variants demonstrate the effects of adapting the physical recipes for different growth channels in the build-up of BH and stellar contents of galaxies. See Appendix C for details.

produced by the MinSeed variant is similar to that produced by the Default variant. The largest values of M_* have already exceeded $10^6 M_\odot$ at $z = 20$, implying that significant growth through the bursty mode has occurred at $z \gtrsim 20$ (see Fig. 14 for some examples). Such a growth is expected to be most significant in the MinSeed variant due to the least effective AGN feedback, thus explaining the similar ranges of M_* produced by the MinSeed and Default variants. The differences among the three variants become less significant when galaxies evolve to higher masses, and vanish at $M_* \gtrsim 10^7 M_\odot$. This is a direct consequence of the self-regulation growth of BHs, as demonstrated by the converged evolution of M_{BH} and M_* in individual galaxies using a series of idealized and controlled experiments done in §5.3.3. The differences in M_{BH} and M_* produced by the seeding methods are thus observable only for extremely low-mass galaxies ($M_* \lesssim 10^7 M_\odot$) or galaxies at very high redshift ($z \gtrsim 10$), for which current observations are still very limited (see §5.4 for a data compilation).

In contrast to the seeding methods, adaptations for the bursty mode can produce larger and longer-lasting effects on the BH and stellar contents of galaxies. This can be seen by comparing the predictions from the three variants with different treatments of nuclear bursts (MinSeed, NoBurst and BurstHighSFE). As discussed in §5.4.2, the growth through the bursty mode plays an essential role in boosting the masses of low-mass BHs to the regime of SMBHs ($M_{\text{BH}} \gtrsim 10^6 M_\odot$). The complete absence of the bursty mode in the NoBurst variant thus produces the lowest M_{BH} at any M_* and any z , and leads to a $M_{\text{BH}}-M_*$ relation that is significantly below the observation results (see §5.4). However, disabling only the wet channel in the bursty mode (BurstHighSFE) yields smaller deficiency in M_{BH} , reinforcing the conclusion reached phenomenologically in Paper-I (see their §5.2 and fig. 7) that moderate boosts of BH masses are sufficient to bring the BHs into the regime where the continuous mode takes over the responsibility of growing BHs to higher masses. Thus, the dry mode alone seems sufficient to grow BHs to the $M_{\text{BH}}-M_*$ relation observed at $z \approx 0$ (panel f). However, at $z \approx 5-10$, the BH masses predicted by the BurstHighSFE variant is more than 0.5 dex lower than that predicted by the MinSeed (or Default) variant at $M_* \lesssim 10^9 M_\odot$, as expected from the inefficiency of the dry channel in growing BHs compared to the wet channel (see §5.3.1). Another interesting feature that can be expected for the BurstHighSFE variant is the peculiar properties of nuclear star clusters (NSCs). This expectation is built on the consideration that AGN feedback dominates the depletion of gas in SNF nuclei in the Default variant, and that little BH accretion during a nuclear burst also implies little competition from the AGN feedback (see §4.5.1). Intense nuclear starbursts are thus inevitable in this variant, which may produce NSCs that are much more massive than those produced in other variants. The presence of such massive NSCs may significantly bias the total light of galaxies towards their centers and produce dot-like morphologies. Fortunately, this regime overlaps with the faint AGN population that has recently been probed by JWST (see §5.4), thus providing a potential way to distinguish the contributions to BH growth through the wet and dry channels in nuclear bursts, and to constrain the coevolution of galaxies, NSCs and BHs. We will come back to this topic in an upcoming paper.

The final variant considered here is AGNConstEff, in which we impose a constant efficiency of AGN feedback at all redshifts (see Eq. A7). Compared with the MinSeed variant, AGNConstEff produces significantly stronger AGN feedback in the continuous mode at $z \gg 0$, and thus prevents the growth of BH and stellar masses in massive galaxies. Noticeable differences can be found at $z \lesssim 5$ when galaxy formation has transited to the continuous era (see §5.3.2). At $z \approx 5$, the upper limit of M_* produced by the AGNConstEff variant

(panel d, orange curve) is about 1 dex lower than that produced by the MinSeed variant (green curve). The self-regulation of BH growth also leads to a reduction in M_{BH} larger than that in M_* , so that the ratio M_{BH}/M_* at given M_* is also lower in the AGNConstEff variant in the high-mass end at $z \lesssim 2$ (panels e and f). The deficit in the number of over-massive galaxies and BHs at $z \approx 5$, and the low values of M_{BH}/M_* at $z \lesssim 2$ suggest that AGNConstEff may not be able to match observations of massive galaxies and BHs, such as high- z quasars and local massive ellipticals. Instead, these observations seem to prefer a redshift-dependent efficiency of the AGN feedback, such as the one adopted in the Default variant.

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