Semi-analytic forecasts for $JWST$ – V. AGN luminosity functions and helium reionization at $z = 2$–7

L. Y. Aaron Yung,[1]⋆† Rachel S. Somerville,[2] Steven L. Finkelstein,[3] Michaela Hirschmann,[4,5] Romeel Davé,[6,7,8] Gergö Popping,[9] Jonathan P. Gardner[1] and Aparna Venkatesan[10]

[1] *Astrophysics Science Division, NASA Goddard Space Flight Center, 8800 Greenbelt Rd, Greenbelt, MD 20771, USA*
[2] *Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave, New York, NY 10010, USA*
[3] *Department of Astronomy, The University of Texas at Austin, Austin, TX 78712, USA*
[4] *DARK, Niels Bohr Institute, University of Copenhagen, Lyngbyvej 2, DK-2100 Copenhagen, Denmark*
[5] *INAF - Astronomical Observatory of Trieste, Via G.B. Tiepolo 11, I-34143 Trieste, Italy*
[6] *Institute for Astronomy, University of Edinburgh, Edinburgh EH9 3HJ, UK*
[7] *University of the Western Cape, Bellville, Cape Town 7535, South Africa*
[8] *South African Astronomical Observatories, Observatory, Cape Town 7925, South Africa*
[9] *European Southern Observatory, Karl-Schwarzschild-Strasse 2, D-85748 Garching, Germany*
[10] *Department of Physics and Astronomy, University of San Francisco, 2130 Fulton Street, San Francisco, CA 94117, USA*

ABSTRACT
Active galactic nuclei (AGN) forming in the early universe are thought to be the primary source of hard ionizing photons contributing to the reionization of intergalactic helium. However, the number density and spectral properties of high-redshift AGN remain largely unconstrained. In this work, we make use of physically-informed models calibrated with a wide variety of available observations to provide estimates for the role of AGN throughout the Epoch of Reionization. We present AGN luminosity functions in various bands between $z = 2$ to 7 predicted by the well-established Santa Cruz semi-analytic model, which includes modelling of black hole accretion and AGN feedback. We then combine the predicted AGN populations with a physical spectral model for self-consistent estimates of ionizing photon production rates, which depend on the mass and accretion rate of the accreting supermassive black hole. We then couple the predicted comoving ionizing emissivity with an analytic model to compute the subsequent reionization history of intergalactic helium and hydrogen. This work demonstrates the potential of coupling physically motivated analytic or semi-analytic techniques to capture multi-scale physical processes across a vast range of scales (here, from AGN accretion disks to cosmological scales). Our physical model predicts an intrinsic ionizing photon budget well above many of the estimates in the literature, meaning that helium reionization can comfortably be accomplished even with a relatively low escape fraction. We also make predictions for the AGN populations that are expected to be detected in future *James Webb Space Telescope* surveys.

Key words: galaxies: active–galaxies: evolution–galaxies: formation–galaxies: high-redshifts–cosmology: theory–dark ages, reionization, first stars

1 INTRODUCTION

Active galactic nuclei (AGN), powered by accreting supermassive black holes (SMBHs), are some of the brightest objects detected in the distant Universe. Driven by a set of physical and radiative processes distinct from the ones that power star-forming galaxies, this class of objects is more efficient at producing hard ionizing photons loosely defined as ionizing radiation with energy sufficient to ionize helium (e.g. >54.4eV). However, these AGN are also relatively rare compared to star-forming galaxies. Furthermore, AGN that are faint or obscured are extremely hard to detect, leaving large uncertainties in their number density and spectral characteristics, which propagate to uncertainties in estimates of the total ionizing photon budget during the Epoch

⋆ E-mail: aaron.yung@nasa.gov
† NPP Fellow

of Reionization (EoR) (e.g. Hopkins et al. 2007a; Madau & Haardt 2015).

The bulk of the relatively bright high-redshift AGN were identified in the large, multi-colour Sloan Digital Sky Survey (SDSS; York et al. 2000; Richards et al. 2002; Jiang et al. 2008; Bovy et al. 2011) with optical fluxes and colour selection techniques. The more recent Baryon Oscillation Spectroscopic Survey (BOSS; Dawson et al. 2013; Ross et al. 2013) based on SDSS-III has identified over 20 000 AGN at $2.2 < z < 3.5$. And at $z \gtrsim 4$, high-redshift AGN can be selected from *Hubble Space Telescope* (*HST*) deep-field surveys, such as the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), with techniques based on rest-frame UV fluxes (Giallongo et al. 2015) or follow-up with X-ray capable instruments, such as the *Chandra* Observatory (Weisskopf et al. 2000; Nandra et al. 2015; Civano et al. 2016; Luo et al. 2016; Xue et al. 2016; Kocevski et al. 2017). The ground-based Subaru Telescope (e.g. Matsuoka et al. 2018, 2019) and the space-based *X-Ray Multi-Mirror Mission* (*XMM-Newton;* Jansen et al. 2001) are also powerful facilities for detecting AGN during EoR. Combining the rich set of available observations has provided significant insights into the number density and distribution of AGN at high redshift, and how they have evolved across cosmic time (Manti et al. 2017; Shen et al. 2020).

The *James Webb Space Telescope* (*JWST*; Gardner et al. 2006) is the next-generation NASA flagship facility, and will be capable of exploring both AGN and galaxies forming in the primordial universe. With its unprecedented infrared (IR) sensitivity, it is expected to detect faint objects several magnitudes below the detection limits of current generation instruments. And with the photometric and spectroscopic capabilities of its on-board instruments, *JWST* is also well-equipped to identify AGN in high-redshift surveys via conventional colour selection techniques and follow-up spectroscopic diagnostics. In addition, the Mid-Infrared Instrument (MIRI) also opens up the possibility of detecting obscured AGN at $z \lesssim 3$ (Yang et al. 2020). Studies have also predicted that *JWST* will be able to provide indirect constraints on the BH masses and accretion rates of AGN, and possibly the seeding and growth mechanisms for BHs in primordial halos (Natarajan et al. 2017; Volonteri et al. 2017; Amarantidis et al. 2019).

Numerical simulations have long been used to investigate and experiment with the physics that drives AGN and their host galaxies. This is especially important for understanding processes that take place on physical scales or time scales that cannot be directly studied with observations. Simulations have been conducted in various flavours, including fully-hydrodynamic, cosmological simulations such as SIMBA (Davé et al. 2019; Thomas et al. 2021), Illustris (Vogelsberger et al. 2014), Illustris-TNG (Weinberger et al. 2017; Nelson et al. 2019), and Horizon-AGN (Dubois et al. 2016; Kaviraj et al. 2017; Volonteri et al. 2016, 2020); 'zoom-in' simulations selected from within a larger cosmological volume (e.g. Choi et al. 2014; Anglés-Alcázar et al. 2017); and idealized, isolated galaxy simulations (e.g. Hopkins & Quataert 2010). These simulations incorporate many of the physical processes expected to be important in shaping BH growth, including mechanical and radiative feedback from SMBHs (e.g. Choi et al. 2014, 2017; Weinberger et al.

2017; Su et al. 2021), although many uncertainties remain about the details of how these processes work. The computational resources required for hydrodynamic simulations increase rapidly with the simulated volume and mass resolution. This tension makes it extremely difficult to simultaneously capture both a large volume and include the full range of relevant physical processes, which take place across a vast range of spatial and temporal scales (e.g. Anglés-Alcázar et al. 2020).

On the other hand, semi-analytic and empirical approaches provide flexible, physically motivated alternatives for efficiently modelling large ensembles of objects (e.g. Fanidakis et al. 2013; Menci et al. 2014; Ricarte & Natarajan 2018a; Dayal et al. 2020; Fontanot et al. 2020; Orofino et al. 2021). Combining the high computational efficiency and the modularized structure of these models, controlled experiments can also be conducted to explore the parameter space and alternative prescriptions for specific physical processes. In analytic or semi-empirical studies that focus on the IGM phase transition, even simpler empirical relations between halos and AGN luminosity are adopted (McQuinn et al. 2009; Madau & Haardt 2015; Hassan et al. 2016; Finkelstein et al. 2019; Faucher-Giguère 2021).

The EoR is one of the most significant events in cosmic history, when the intergalactic medium (IGM) transitioned from neutral to ionized (Fan et al. 2006a,b). Although it is unlikely that AGN contributed significantly during hydrogen reionization, AGN are thought to be the main contributor to the reionization of helium, which could have begun as early as $z \sim 7$ (e.g. Madau & Haardt 2015; Finkelstein et al. 2019), and does not conclude until $z \lesssim 3$, as inferred by the He II Gunn-Peterson effect observed in quasar spectra (e.g. Jakobsen et al. 1994; Heap et al. 2000; Syphers & Shull 2013).

Currently, large uncertainties remain in the ionizing photon budget contributed by AGN at high redshift. This can be broken down into three moving parts: the number density of sources, their spectral characteristics, and the fraction of ionizing photons that were able to escape to the IGM. For the relatively bright AGN, the first two have been quite well constrained with the aforementioned photometric and spectroscopic surveys. However, both of these remain unconstrained for the fainter populations, which are potentially the main contributors that drove the reionization process at $z \gtrsim 4$.

The escape fraction of ionizing photons is the least constrained among the three components and is also notoriously difficult to model. Simulations have shown that this quantity is highly sensitive to a large set of intricate, multi-scale geometrical and physical features, including mass and angular momentum of galaxies (e.g. Paardekooper et al. 2011), density profile and distribution of sources (e.g. Benson et al. 2013; Kimm et al. 2019; Ma et al. 2020), gas and dust content (e.g. Popping et al. 2017), and stellar feedback effects (e.g. Kimm & Cen 2014; Kimm et al. 2017; Trebitsch et al. 2017). Many studies attempting to constrain the escape fraction via observations have arrived at similar conclusions (e.g. Dijkstra et al. 2016; Guaita et al. 2016; Shapley et al. 2016; Fletcher et al. 2019; Nakajima et al. 2020). Although these studies mostly focus on star-forming galaxies, these complications are expected to also apply to AGN, with the added complication that photons must propagate out of the op-

tically thick material that fuels the black hole (torus) as well as through the interstellar medium (ISM) of the host galaxy and circumgalactic medium (CGM). Studies have also shown that the escape fraction can easily have several orders of magnitude of scatter and does not correlate well with any particular global galaxy property (e.g. Ma et al. 2015; Paardekooper et al. 2015).

In this work, we construct a physical, source-driven modelling pipeline that integrates many multi-scale physical processes, ranging from accretion activity near BHs to halo-scale gas accretion and cooling, and explore their collective effects on cosmological-scale processes such as the progression of intergalactic helium reionization. The models are calibrated to reproduce low-redshift observations, and are then used to make predictions for AGN populations that are too faint and at too high a redshift to have existing direct observational constraints. Our predicted hard ionizing photon emissivity accounts for the combined effects of the cosmological physics-based model for the AGN population and a spectral model that reflects the underlying BH masses and accretion rates. By interfacing these model components to an analytic reionization model for intergalactic hydrogen and helium (e.g. Madau et al. 1999), we can make inferences about the ionizing escape fraction required to match the existing constraints. The modelling pipeline developed in this work is an extension of the versatile semi-analytic model for galaxy formation, which tracks a wide variety of galaxy formation physics and allows some of our novel predictions to be self-consistently compared to the properties of their host galaxies (Somerville & Primack 1999; Somerville et al. 2008; Hirschmann et al. 2012).

In this series of *Semi-analytic forecasts for JWST* papers, we present a comprehensive collection of predictions for galaxies and AGN forming in the early universe that are anticipated to be observed by *JWST* and other future facilities. In Yung et al. 2019a,b (hereafter Paper I and Paper II), we presented detailed predictions for the photometric and physical properties for high-redshift galaxy populations, for which we also provided predictions for how future *JWST* detections can be used to further constrain galaxy formation physics. In Yung et al. 2020a,b (hereafter Paper III and Paper IV), we modelled the production of ionizing photons by stars in these galaxies and further investigated the subsequent reionization history of intergalactic hydrogen. Results from these previous works paint a coherent picture showing that the predicted galaxy population, which is able to reproduce a wide range of observed distributions of $M_{\rm UV}$, M_*, and SFR up to $z \sim 10$, are also producing sufficient ionizing photons to yield a reionization history that is consistent with all of the existing IGM and CMB constraints. In this penultimate paper of the series (Paper V), we extend our predictions to include AGN and their contributions to the reionization of intergalactic hydrogen and helium. All results presented in the paper series will be made available at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/. Full object catalogues will be released as part of an upcoming, final paper of this series (Yung et al., in preparation; or Paper VI).

The key components of this work are summarized as follows: the key modelling components are summarized briefly in Section 2. Predicted AGN characteristics and the resulting helium reionization history are presented in Section 3.

We also investigate the role AGN played during cosmic hydrogen reionization in Section 4. We discuss our findings in Section 5, and a summary and conclusions follow in Section 6.

2 THE MODELLING FRAMEWORK

In this section, we present the components that make up the fully semi-analytic modelling pipeline for AGN growth and cosmic He reionization. Throughout this work, we adopt cosmological parameters that are consistent with the ones reported by Planck Collaboration (XIII 2016a): $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8$ km s^{-1}Mpc^{-1}, $\sigma_8 = 0.831$, and $n_s = 0.9665$. We adopt hydrogen and helium mass fractions $X = 0.75$ and $Y = 0.25$, respectively.

2.1 Semi-analytic model for galaxies and AGN

The foundation of this series of papers is the Santa Cruz semi-analytic model (SAM). This modelling framework is described in full in Somerville & Primack (1999), Somerville, Primack & Faber (2001), Somerville et al. (2012), Popping et al. (2014, hereafter PST14), Somerville et al. (2015, hereafter SPT15), and in particular, the co-evolution model for galaxies, black holes and AGN in Somerville et al. (2008, hereafter S08), Hirschmann et al. (2012), and Porter et al. (2014). The free model parameters are calibrated to a subset of $z \sim 0$ observations without recalibration for higher redshifts. See appendix B in Paper I and section 2.2.3 in Somerville et al. (2021) for details regarding the calibration criteria and process.

Dark matter halo merger and growth histories, or simply 'merger trees', are the backbone of the semi-analytic modelling approach. These merger trees can either be extracted from cosmological-scale numerical simulations or constructed using the extended Press-Schechter (EPS) formalism (Press & Schechter 1974; Bond et al. 1991; Sheth & Tormen 2002). Although numerical merger trees are better at capturing the interplay between halos and the cosmic environment in which they are embedded, these simulations are subject to tension between the simulated volume and resolution, and can be computationally expensive. Both a large simulated volume and high mass resolution are required to properly capture the black hole seeding in early, low-mass halos, while sampling rare massive halos with their merger histories sufficiently resolved. On the other hand, EPS-based algorithms are able to flexibly generate merger trees for halos of any given masses *on-demand*, which have been shown to reproduce the statistical results for a large ensemble of merger trees extracted from N-body simulations (Lacey & Cole 1993; Somerville & Kolatt 1999; Zhang et al. 2008; Jiang & van den Bosch 2014).

In this work, we adopt the EPS merger tree algorithm presented in Somerville & Kolatt (1999) and S08. For each output redshift, we create a grid of 200 halo masses that are equally spaced in the range $V_{\rm vir} = 100 - 1400$ km s^{-1}. Note that the halo mass range used in this work is adjusted from previous papers in the series to better capture the rare, massive halos that are likely to host SMBH. For each of the masses in the grid, we apply the algorithm to construct a hundred Monte Carlo merger history realizations. These

merger histories trace back to either a minimum progenitor mass of $M_{\rm res} = 10^{10} {\rm M}_\odot$ or 1/100th of the root halo mass, whichever is smaller. Independent output halo mass grids are created between $z = 2$ to 7 at half redshift increments. The expected number density of each of these dark matter halos is weighted by the fitted halo mass function in the respective redshift provided by Rodríguez-Puebla et al. (2016) based on results from the MultiDark simulation suite (Klypin et al. 2016). See appendix C in Paper I for details regarding the halo mass function adopted for this work.

The versatile Santa Cruz SAM is well-equipped to capture the symbiotic relationship between AGN and their host galaxies, where the cold gas reservoir that fuels the accreting SMBH is constantly modified by a large collection of cooling and galaxy formation physics, and in turn, the star formation activity in the host galaxy is regulated by AGN feedback. We refer the reader to Paper I for a full description of the galaxy formation model configurations adopted in this work and the rest of the series. Here we provide a concise summary of the model components that are related to the seeding and growth of SMBHs only.

2.1.1   The growth of bulges

In the Santa Cruz SAM, a basic ansatz is that BH properties are tightly connected to the properties of the *bulge* component of galaxies, as supported by observations. When gas initially cools, it is assumed to accrete into a disc, and stars that form out of that gas are assumed to have disc-like kinematics and morphology. Disc stars can be moved into a bulge component via two mechanisms. The first is mergers: when galaxies merge, if the merger mass ratio is larger than a critical value, all stars from both progenitors are placed into a bulge component. For lower mass ratio mergers, the stars from the lower mass progenitor are deposited in the bulge component of the descendent galaxy. This approach is motived by results from numerical simulations of galaxy mergers and is widely used in semi-analytic models. We refer the reader to S08 and Hirschmann et al. (2012) for the relevant references and further details.

The second way for bulges to grow is via 'disc instabilities'. A 'disc instability' (DI) mode for bulge formation and growth has been implemented in the Santa Cruz SAM by Hirschmann et al. (2012); Porter et al. (2014), and a handful of DI model variants have been implemented and tested in these works. Here we provide a concise summary of the DI model adopted in this work, which has been shown to be in good agreement with observed galaxy morphologies from $z \sim 2$–0 (Brennan et al. 2015), and we refer the reader to the above references for a full description. Following early analytic models by Toomre (1964) and Efstathiou, Lake & Negroponte (1982), a disc becomes unstable when the ratio of dark matter mass to disc mass falls below a critical value

$$M_{\rm disc,crit} \equiv \frac{V_{\rm max}^2 r_{\rm disc}}{G \epsilon_{\rm crit}}, \tag{1}$$

where $V_{\rm max}$ is the maximum circular velocity of the halo, $r_{\rm disc}$ is the scale length of the cold gas disc, and $M_{\rm disc}$ is the combined mass of stars and gas in the disc. Therefore, by defining a disc coefficient as follows

$$\epsilon_{\rm disc} \equiv \frac{V_{\rm max}}{\sqrt{G M_{\rm disc}/r_{\rm disc}}}, \tag{2}$$

a disc becomes unstable when $\epsilon_{\rm disc} < \epsilon_{\rm crit}$. Numerical simulations report a range of $\epsilon_{\rm crit} \sim 0.6$ to 1.1 depending on the composition of the disc, where discs with higher gas content tend to have a lower instability threshold than the ones with lower gas content. We adopt $\epsilon_{\rm crit} = 0.75$, which is chosen to reproduce the observed bulge fractions as a function of mass in nearby galaxies. At each time-step, if the disc is unstable, a fraction of stars and gas is moved from the disc to the bulge to achieve marginal stability, with the cold gas assumed to fuel and be consumed by a starburst. The fraction of gas and stars being moved is proportional to the ratio of cold gas and stars in the disc.

2.1.2   BH seeding and growth

A seed black hole of fixed mass $M_{\rm seed}$ is assigned to every 'top-level' halo in our merger trees. In our fiducial configuration, we adopt a BH seed mass of $M_{\rm seed} = 10^4 {\rm M}_\odot$. This is somewhat larger than the mass of a seed BH that would be expected to be left behind by massive Population III stars (Abel et al. 2002), but our merger trees do not reach the halo masses that are expected to host these Pop III stars, so this accounts for some previous growth. Previous testing has shown that most results are not sensitive to BH seed mass values within the range of $M_{\rm seed} \sim 100 {\rm M}_\odot$ to $10^4 {\rm M}_\odot$ (see S08).

Rapid BH growth results in the radiatively efficient 'bright mode' of AGN activity. This mode is fuelled by cold gas accretion, usually triggered by mergers and disc instabilities (e.g. Hopkins et al. 2007c). In the Santa Cruz SAM, radiatively efficient accretion is triggered when a merger with mass ratio of $\mu > 0.1$ occurs, where μ is the mass ratio of the baryonic components and the dark matter within the central part of the galaxy (see S08 for the precise definition). The BH in the two progenitor galaxies are assumed to merge rapidly to form a new BH, with mass conserved. Motivated by results from simulations presented in Hopkins et al. (2007b), the post-merger BH will then grow at the Eddington rate until it reaches a critical mass

$$\log \left(\frac{M_{\rm BH,crit}}{M_{*,\rm bulge}} \right) = f_{\rm BH,crit} [-3.27 + 0.36 \, {\rm erf}((f_{\rm gas} - 0.4)/0.28)], \tag{3}$$

where $M_{*,\rm bulge}$ is the stellar mass of the bulge and $f_{\rm gas}$ is the cold gas fraction of the larger progenitor. $f_{\rm BH,crit}$ is a free parameter that is calibrated such that the output $M_{\rm bulge}$–$M_{\rm BH}$ relation reproduces the relation observed at $z \sim 0$ (McConnell & Ma 2013). This critical mass corresponds to the energy needed to halt further accretion and begin to power a pressure-drive outflow. After reaching the critical BH mass, the accretion rate declines as a power law as described in (Hirschmann et al. 2012).

We also specify an assumed Gaussian scatter in the BH critical mass, which is set to $\sigma_{\rm BH} = 0.3$ in our fiducial model (e.g. Somerville et al. 2021). In addition, we also include an adjusted model with $\sigma_{\rm BH} = 0.5$, which has been shown in past studies to better reproduce the bright AGN population at high redshift (Hirschmann et al. 2012). We note that
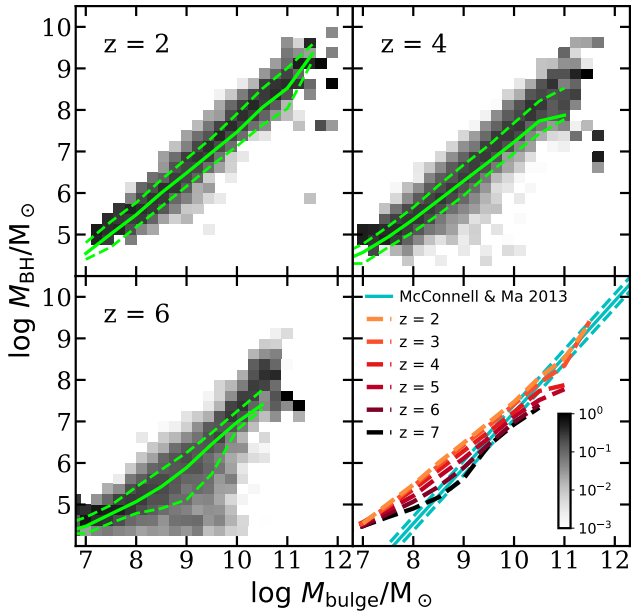
Figure 1. The $M_{\rm BH}$–$M_{\rm Bulge}$ relation with $\sigma_{\rm BH} = 0.3$ predicted by our fiducial model at $z = 2$, 4, and 6. The greyscale 2D histograms show the conditional number density per Mpc^3, which is normalized to the sum of the number density in its corresponding (vertical) $M_{\rm Bulge}$ mass bin. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The last panel shows an overlay of the median relation from $z = 2$ to 7. This is compared to the observed constraints from McConnell & Ma (2013) at $z = 0$, which was used to calibrate our model.

the intrinsic scatter in the $M_{\rm bulge}$–$M_{\rm BH}$ relation at $z = 0$ is approximately 0.3 dex, and it could have been larger at high redshift, as mergers and other processes will reduce the scatter and the scatter is essentially unconstrained at high redshift. Later in this work, we will also show that the additional scatter in the relation is one way to help produce more bright AGN and yield better agreement with the observed number density for the bright AGN populations.

Rapid BH growth can also follow a disc instability, where we allow a fraction of the mass of the disc $f_{\rm BH,disc} = 0.001$ to accrete onto the BH. This additional fuel is available to be accreted by the BH after a DI event until it is fully consumed. The DI-associated fueling is not required to respect the critical mass associated with merger-driven fueling, however the timescale for the BH accretion (lightcurve) is given by the same parameters used to define the lightcurve for merger-driven fueling. Given both the higher gas fractions (e.g. Paper II) and gas inflow rates at high redshifts, DI events occur quite frequently and are a major channel for BH growth at early times. This model is able to reproduce both the observed AGN bolometric luminosity functions and the BH mass–bulge relation observed at $z \sim 0$. Fig. 1 shows our model predictions at $z = 2$ to 7 for the $M_{\rm BH}$–$M_{\rm bulge}$ relation. The additional scatter to low BH masses at $z \gtrsim 6$ is likely a result of the lack of time for early forming BH to grow and reach the target relation, given the short age of the Universe at this redshift.

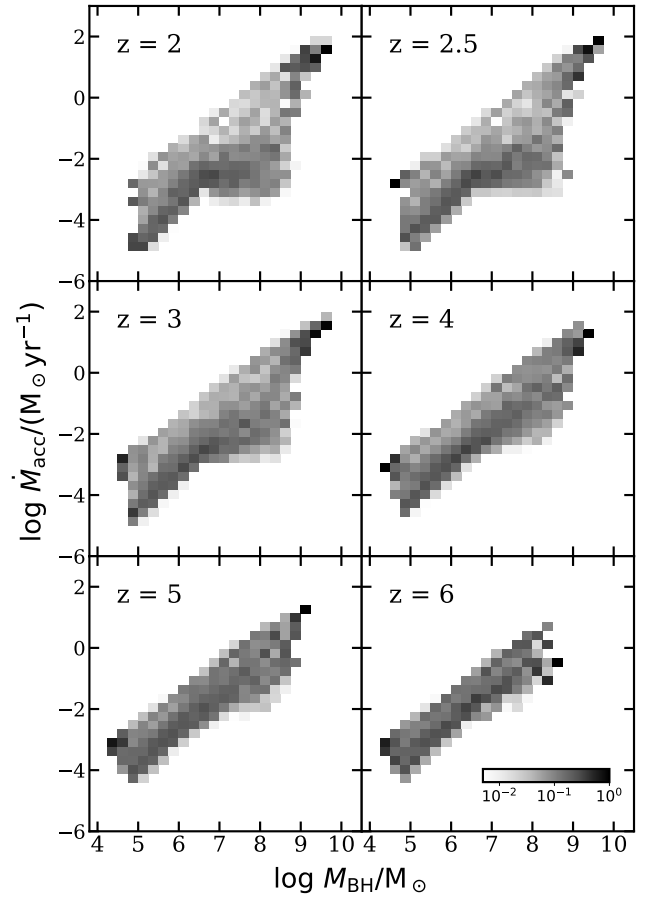A secondary, radiatively inefficient 'radio' mode driven by Bondi-Hoyle accretion (Bondi 1952) in hot quasi-

Figure 2. Predicted conditional distribution of black hole mass, $M_{\rm BH}$, versus BH accretion rate, $\dot{M}_{\rm acc}$, between $z = 2$ to 6. The 2D histograms are colour-coded for the conditional number density per Mpc^3, which is normalized to the sum of the number density in its corresponding (vertical) $M_{\rm BH}$ bin.

hydrostatic halos with the isothermal cooling flow model proposed by Nulsen & Fabian (2000) is also included. Both of these accretion modes have been accounted for in the total accretion rate $\dot{M}_{\rm acc}$ and the subsequent feedback effects on the ISM and star formation models. However, the contribution of this mode to the BH accretion rates, especially at high redshift, is negligible. As reported in Paper I, the feedback effects of this accretion mode on star formation are also insignificant at $z \gtrsim 4$.

Fig. 2 shows the predicted redshift evolution of the joint distribution of $\dot{M}_{\rm acc}$ and $M_{\rm BH}$ at $z = 2$ to 6. The top boundary of the distribution corresponds to the Eddington accretion rate, as our model does not include physical processes that permit accretion at super-Eddington rates. The distribution of accretion rates shifts to lower values and the scatter increases towards low redshift with the decline of the merger rate, galaxy gas fractions, and the incidence of disc instabilities.

2.2 A physical AGN spectral model

In this work, we adopt the physical AGN spectral model developed by Kubota & Done (2018, hereafter KD18; also

see Done, Gierliński & Kubota 2007 and Done et al. 2012).
This model takes BH masses and accretion rates as inputs
and produces physically-motivated SEDs spanning a wide
energy range. The KD18 work separately models the emission from three distinct radiative 'zones' near an accreting
SMBH. These zones include a spherical hot Comptonization
region and a standard disc consisting of a warm Comptonization region and a cooler outer disk region. For each of these
regions, the emission is modelled assuming blackbody emissivity as described by Novikov & Thorne (1973). In addition,
this model accounts for the geometrical self-heating effect
due to radiation originating from the hot Comptonization
region that subsequently illuminates and reheats the warm
Comptonization and outer disc regions. This effect is labelled
as 'hard X-ray reprocess' in the KD18 model.

The QSOSED model form KD18 takes the dimensionless
accretion rate defined as $\dot{m} \equiv \dot{M}_{\mathrm{acc}}/\dot{M}_{\mathrm{edd}}$, where $\dot{M}_{\mathrm{acc}}$ is the
black hole accretion rate predicted and $\dot{M}_{\mathrm{edd}}$ is the Eddington accretion rate calculated internally in the SAM based
on the mass of the BH. For AGN predicted in this modelling pipeline, these quantities are passed to QSOSED from
the Santa Cruz SAM. For the rest of the free parameters, we
follow the default configuration from KD18 that is guided
by observed spectra of individual nearby AGN. These free
parameters correspond to electron temperature (kT_e), spectral index (Γ), and outer radius (R) for the hot and warm
Componization components, and a handful of others parameters that characterize the scale height and radii of the disc
component (see table 3 in KD18). The calibration of these
free parameters is guided by observed high-resolution AGN
spectroscopy measurements (Mehdipour et al. 2011, 2015;
Jin et al. 2012b,a). Note that these spectra are either corrected for galactic interstellar reddening or do not suffer from
strong reddening effect in the first place. For this reason,
dust attenuation and obscuration effects will be modelled
separately when these SEDs are being forward-modelled to
observable quantities (e.g. UV1450 luminosities).

In Fig. 3, we show a range of sample output AGN spectra from QSOSED for a range of fixed values of $\dot{m}$ and M_{BH}
computed for an energy grid spanning 10^{-5} to 10^3 keV. We
show a case where we explore the sensitivity of the SED
model to BH accretion rate by showing outputs for a range
of $0 > \log \dot{m} > -1.67$ for a fixed $\log(m_{\mathrm{BH}}/\mathrm{M}_\odot) = 8$. Similarly, we also show a similar experiment with BH masses by
showing outputs for a range of $5 < \log M_{\mathrm{BH}}/\mathrm{M}_\odot < 10$ for a
fixed $\dot{m} = -0.5$. Furthermore, QSOSED also takes inclination
of the accretion disc as a free parameter, on which UV emission from the disc has a strong dependence, as shown in the
bottom panel of Fig. 3. For the rest of this work, we adopt
a fixed inclination of 45° for an overall averaged, isotropic
SED for emission quantities such as the emissivity of ionizing photons. On the other hand, for observable quantities,
such as band luminosities, we randomly assign an inclination of 0° < i < 90° to account for the possible distribution
of inclinations along our line-of-sight. We note that the inclination of the disc has very little effect in the energy range
that corresponds to that of hard ionizing radiation, which
originates from the Hot and Warm Comptonization regions.

For comparison, we add two cases of power-law spectra, $L_\nu = L_{912}(\nu/\nu_{912})^{\alpha_\nu}$, adopted by similar studies. We
include a single power law spectrum with $\alpha_\nu = -1.57$ as
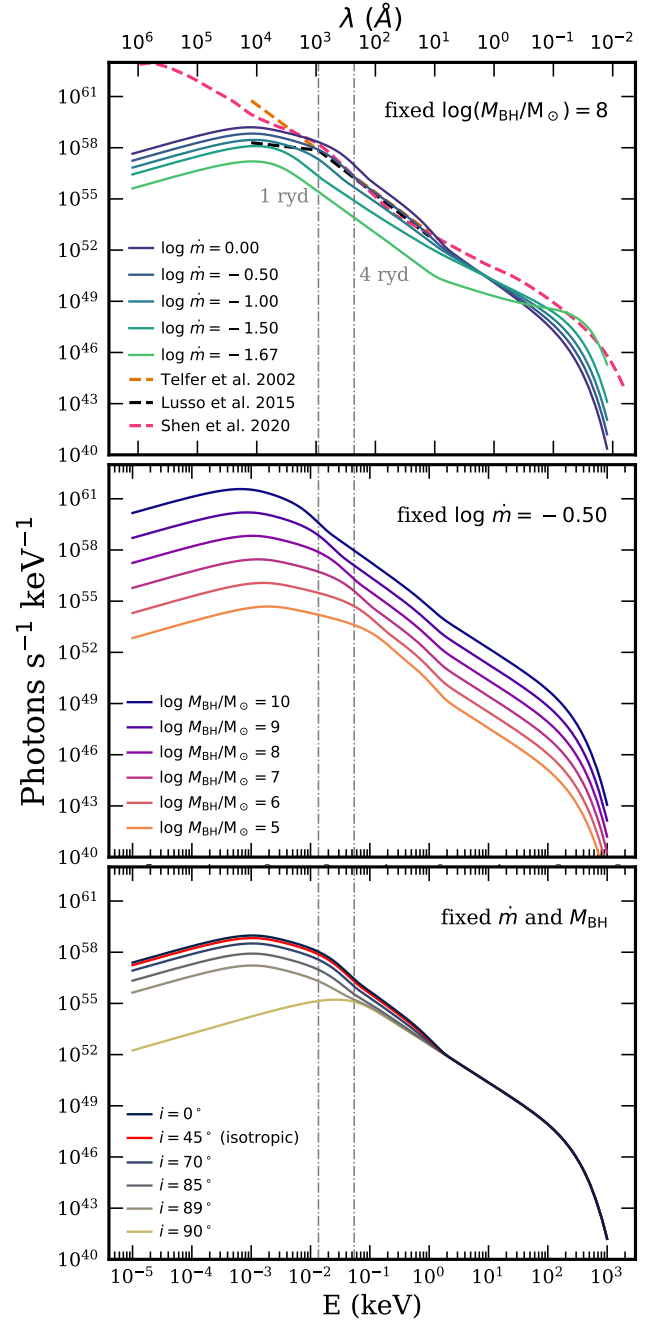reported by Telfer et al. (2002), as adopted in Hassan et al.

Figure 3. Rest-frame AGN SEDs predicted with QSOSED for a
range of input Eddington-normalized black hole accretion rates,
$\dot{m}$, black hole masses, M_{BH}, and inclinations, i. In addition, we
also show the wavelength in angstrom corresponding to the energy (top axis). The *top panel* shows predictions from a range
of $\log \dot{m} = [0, -1.67]$ assuming a fixed $m_{\mathrm{BH}} = 10^8$ M$_\odot$; and the
middle panel shows SEDs for a range of $\log \dot{M}_{\mathrm{BH}}/\mathrm{M}_\odot = [5, 10]$
assuming a fixed $\dot{m} = -0.5$. The *bottom panel* shows a range of
inclination $i = [0°, 90°]$ for a fixed $\dot{m}$ and M_{BH}. We also mark the
energy levels at 1 and 4 ryd to guide the eye. We also show examples of a broken power law with $\alpha_\nu = -1.7$ and -0.61 for $\lambda \leq 912$
and $\lambda > 912$, respectively (Lusso et al. 2015, black dashed line);
a simple power law for $\alpha_\nu = -1.57$ (Telfer et al. 2002, orange
dashed line); and the mean SED template adopted in Shen et al.
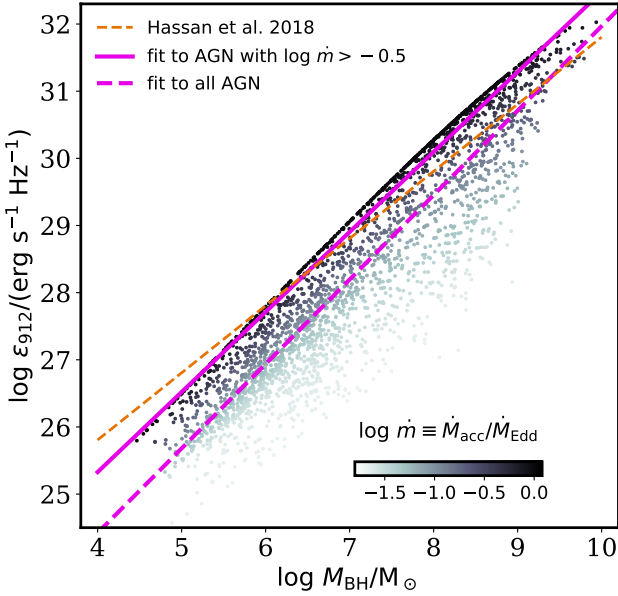(2020). See text for details.

Figure 4. Distribution of M_{BH} versus specific UV emissivity at 1 ryd for our simulated AGN populations, colour-coded by the Eddington-normalized accretion rate. Here we show AGN from an output at $z = 3$; however, this relation does not explicitly depend on redshift. The solid magenta lines show the scaling relation fitted to a subset of AGN with high accretion rates $\log \dot{m} > -0.5$, and the dashed magenta line is the fit to all AGN. The orange dashed line shows the empirical relation adopted in Hassan et al. (2018). See text for details.

(2018) and Haardt & Madau (2012), and a broken power law reported by Lusso et al. (2015) with $\alpha_\nu = -1.7$ for at $\lambda \leq 912\text{Å}$ and $\alpha_\nu = -0.61$, as adopted in the Shen et al. (2020) and Madau & Haardt (2015) ionizing emissivity calculation. Both of these assume a black hole mass of $m_{BH} = 10^8 M_\odot$ and are converted to L_{912} based on the prescriptions used in Hassan et al. (2018). We also include a composite SED adopted by Shen et al. (2020), which is normalized to $\nu L_\nu \approx 45.5$ erg s^{-1} at 2500Å. We note that the deviation from the Shen et al. model in the low energy range is due to the lack of dust emission in the KD18 spectral model.

In Fig. 4, we illustrate the scatter in the specific emissivity at 1 ryd, ϵ_{912}, for the AGN populations predicted by the Santa Cruz SAM when coupled with QSOSED. Here we show the $z = 3$ simulated population as an example, but these quantities do not explicitly depend on redshift. We add a case where ϵ_{912} is calculated based on black hole masses using a set of empirical conversions, similar to the procedure used in Hassan et al. (2018), which adopts relations between M_{BH} to B-band luminosity L_B and between L_B to L_{912} (Schirber & Bullock 2003; Choudhury & Ferrara 2005). It is important to note that while the calculation with simple power-laws is in broad agreement with predictions from a physical spectral model, such an approach doesn't capture the scatter introduced by black hole accretion rate, which is shown not to be correlated with black hole mass. Here we provide a fit for this relation that takes the form of $\log \epsilon_{912} = A \log M_{BH} + B$ fitted with a non-linear least

square method. For the high accretion rate population with $\log \dot{m} > -0.5$, we find best-fit parameters $A = 1.19$ and $B = 20.56$, which is fairly similar to the empirical relation derived from observations of luminous AGN. However, we find that the inclusion of lower accretion rate AGN populations shifts the relation down to $A = 1.26$ and $B = 19.39$. Thus empirical relations derived from observations may misrepresent the full population of AGN due to possible selection biases.

We also note that with the default dissipated energy from the hot Comptonization region set at $L_{\mathrm{diss,hot}} = 0.02\, L_{\mathrm{edd}}$, the spectral index for the hot Comptonization region, Γ_{hot}, decreases rapidly for $\log \dot{m} \lesssim -1.7$, which becomes incompatible with the disc-corona geometry assumed by the KD18 model. Given that observational evidence has suggested that the physical properties of these accreting AGN can change drastically across some accretion rate threshold, leading to a drastic drop in radiative efficiency (e.g. Trump et al. 2011), for the rest of this work, AGN with $\log \dot{m} \lesssim -1.7$ are assumed to be radiatively inefficient. While the SAM does predict that a population of such low accretion rate BHs exists, we exclude these objects from spectral modelling and their contribution to luminosity functions and ionizing photon budgets. We refer the reader to Hirschmann et al. (2014) for a more in-depth discussion of radiatively inefficient BHs.

The QSOSED model is accessed through the spectral fitting package XSPEC[1] (Arnaud 1996). Compton emission at each annulus is carried out with the NTHCOMP algorithm (Zdziarski et al. 1996; Zycki et al. 1999) provided as part of XSPEC. The original code in `fortran` is ported to work in a `python` environment with the `f2py` tool (Peterson 2009). Other miscellaneous calculations are carried out using functions from `astropy` (Robitaille et al. 2013; Price-Whelan et al. 2018), `numpy` (van der Walt et al. 2011), and `scipy` (Virtanen et al. 2020).

2.3 Analytic model for cosmic helium reionization

In this section, we present the set of equations that comprise the analytic model that tracks the cosmic helium reionization history. As discussed in Wyithe & Loeb (2003) and Haardt & Madau (2012), even though the ionization front of singly ionized helium may briefly overtake that of hydrogen at very early times (e.g. $z \sim 10$), the expansion of the hydrogen ionization front is not inhibited by the presence of helium simply because most of the ionizing photons produced are only able to ionize hydrogen (< 24.6 eV). Therefore, the following restrictions, $Q_{\mathrm{He\,III}} \leq Q_{\mathrm{H\,II}}$ and $Q_{\mathrm{He\,II}} \leq Q_{\mathrm{H\,II}} - Q_{\mathrm{He\,III}}$, are valid and are often assumed in analytic models. This allows us to model the reionization of intergalactic helium independently of the hydrogen reionization.

Taking an approach similar to that described in Paper III, we calculate the production rate of helium ionizing photons from AGN by integrating the physically modelled spectra. This approach has been shown to be better at capturing subtle changes in the physical properties of the underlying sources, in this case BH mass and accretion rate distri-

---

butions of the AGN population. This approach also provides a realistic estimate of scatter in the AGN properties that is not captured with empirical or power-law conversions. The production rate of He II ionizing photons, $\dot{N}_{\rm ion,He\,II}$, from an individual AGN is calculated by integrating its SED at energies above 54.4 eV ($\lambda < 228$ Å)

$$\dot{N}_{\rm ion,He\,II} = \int_{\nu_{228}}^{\infty} L_{\nu,\rm AGN}(h\nu)^{-1} d\nu. \qquad (4)$$

Here, $L_{\nu,\rm AGN}$ is the AGN spectrum from the QSOSED model presented (see Section 2.2), based on $M_{\rm BH}$ and $\dot{m}$ predicted by the Santa Cruz SAM (see Section 2.1). Similar to Fig. 4, in Fig. 5 we show the scaling relation of $L_{\rm bol}$ versus $\dot{N}_{\rm ion,He\,II}$ for the predicted AGN populations at $z = 3$, as the model components involved in this calculation do not have any explicit redshift dependence. We show that bolometric luminosity is a better tracer for $\dot{N}_{\rm ion,He\,II}$, where the relation is relatively tight. Scatter in this relation arises mainly from differences in the spectrum in the UV, which is dominated by the dependence on $\dot{m}$ (see top panel in Fig. 3). Fitting these data points with a non-linear least square method with a functional form $\log \dot{N}_{\rm ion,He\,II} = A \log(L_{\rm bol}) + B$, we find the best-fit parameters for $A = 0.87$ and $B = 44.27$ for the bright AGN populations with $\log L_{\rm bol}/\rm L_{\odot} > 12$, which extends into the faint AGN and reproduces well the population accreting near the Eddington rate. We also show the case for $\log L_{\rm bol}/\rm L_{\odot} < 12$, where the relation is slightly shifted, best-fitted with $A = 1.02$ and $B = 42.39$. This steeper relation is caused by the additional scatter in the faint population from AGN with a range of BH accretion rate.

Combining the predicted AGN number density and their spectral properties, we calculate the comoving He II ionizing emissivity, $\dot{n}_{\rm ion,He\,II}$, by summing the contribution across the predicted AGN populations at each output redshifts

$$\dot{n}_{\rm ion,He\,II} = \sum_i n_{{\rm h},i}\,\dot{N}_{\rm ion,He\,II},_i\,f_{\rm esc,He},_i. \qquad (5)$$

Here, $n_{\rm h}$ is the number density per Mpc^3 for each AGN i, assigned based on the virial mass of the host halo. Last, $f_{\rm esc,He}$ is the escape fraction of helium ionizing photons specifically for AGN. For the remainder of this work, we refer to the AGN-specific escape fraction as $f_{\rm esc}$ unless otherwise specified. For the rest of this work, we refer to this quantity as the ionizing photon production rate when assuming $f_{\rm esc}=$ 1.00, which is to be distinguished from the emissivity that accounts for the amount of ionizing radiation trapped or dissipated in the ISM and CGM, as well as the the potential dust and neutral gas obscuration occurring in the vicinity of the AGN (e.g. nuclear- or torus-scale obscuration). We also note that the physical processes governing the escape fraction of radiation from AGN can be quite different from the ones affecting the escape fraction of radiation from stellar populations.

The comoving He II ionizing emissivity is then forward modelled into the *volume-averaged* ionizing volume-filling fraction of doubly ionized helium, $Q_{\rm He\,III}$, with the temporal evolution described by the following first-order differential equation

$$\frac{dQ_{\rm He\,III}}{dt} = \frac{\dot{n}_{\rm ion,He\,II}}{\bar{n}_{\rm He}} - \frac{Q_{\rm He\,III}}{\bar{t}_{\rm rec,He\,III}}, \qquad (6)$$

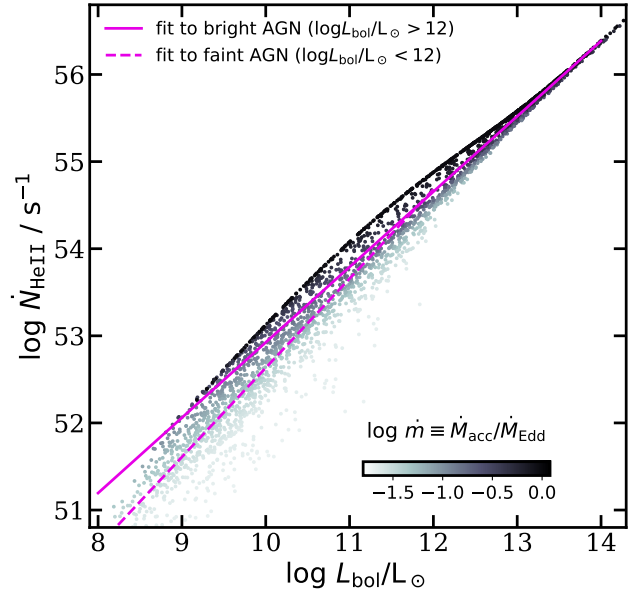as derived in Madau et al. (1999). The two terms separately

Figure 5. Distribution of $L_{\rm bol}$ versus $\dot{N}_{\rm ion,He\,II}$ for our simulated AGN population. Here we show AGN from at output at $z = 3$; however, this relation does not explicitly depend on redshift, colour-coded by the Eddington-normalized accretion rate. The solid magenta line shows the scaling relation fitted only to the bright AGN population with $\log L_{\rm bol}/\rm L_{\odot} > 12$, and is plotted as an extrapolation below this limit. The dashed magenta line is the fit to AGN with $\log L_{\rm bol}/\rm L_{\odot} < 12$. This illustrates how adopting relations based on the brightest AGN may over-predict the mean ionizing photon production rate for the full AGN population.

account for the growth of ionized volume and the sink of ionized intergalactic He III due to recombination. The growth term is the ratio of the comoving He II ionizing emissivity, $\dot{n}_{\rm ion,He\,II}$, and the volume averaged comoving number density of intergalactic helium, $\bar{n}_{\rm He}$. Here, we adopt IGM mean hydrogen density $\bar{n}_{\rm H} = 1.9 \times 10^{-7}$ cm^{-3} (Madau & Dickinson 2014) and the conversion $\bar{n}_{\rm He} = \bar{n}_{\rm H} Y/4(1-Y) = 1.58 \times 10^{-8}$ cm^{-3}. The sink term is characterized by the recombination of He III with free electrons by taking the ratio of $Q_{\rm He\,III}$ and the recombination time-scale of He III. The recombination time-scale of intergalactic helium is given by

$$\bar{t}_{\rm rec,He\,III} = [C_{\rm He\,III}\,\alpha_{\rm B,He\,II}(T/Z^2)\,(\bar{n}_{\rm H} + 2\bar{n}_{\rm He})\,(1+z)^3\,Z]^{-1}, \qquad (7)$$

where $\alpha_{\rm B,He\,III}(T)$ is the case B recombination coefficient for He III (Hui & Gnedin 1997) and $C_{\rm He\,III}$ is the He III clumping factor. In this work, we follow Finkelstein et al. (2019) in assuming an IGM temperature of $T = 2 \times 10^4$ K and $C_{\rm He\,III} = C_{\rm H\,II}$. For $C_{\rm H\,II}$, we adopt the redshift-dependent clumping factor from the radiation-hydrodynamical simulation L25N512 by Pawlik et al. (2015), in which $C_{\rm H\,II}$ evolves from ~ 1.5 to ~ 4.8 between $z \sim 14$ to $z \sim 6$. At $z \lesssim 6$, the clumping factor is extrapolated using a polynomial fit such that the clumping factor continue to increase approximately linearly to a value of ~ 6.8 at $z \sim 2$. The helium reionization history, $Q_{\rm He\,III}(z)$, is then obtained by solving equation (6) using `scipy.integrate.odeint` and as-

`tropy.cosmology` (Robitaille et al. 2013; Price-Whelan et al. 2018).

3 RESULTS

In this section, we present the predictions of our fully semi-analytic, source-driven modelling pipeline. Our results are organized in two parts. (1) the predicted bolometric and band-specific luminosity functions compared to observational constraints and simulated results from numerical hydrodynamic simulations. (2) The ionizing photon production rate for the predicted AGN populations and the subsequent implications for the cosmic helium reionization history.

3.1 AGN luminosity functions

One-point distribution functions provide an effective overview for the statistical characteristics of specific observable or physical property among a large ensemble of objects. The distribution functions for luminosity in specific bands are commonly referred to as luminosity functions (LFs).

3.1.1 Hard X-ray luminosity function

In Fig. 6, we show the hard X-ray luminosity functions (HXLFs) at $z = 2$–7, where the rest-frame hard X-ray luminosity, L_X, is obtained by integrating the AGN spectra between 2–10 keV. We consider the hard X-ray LF as the most robust calibration for our models, as we can fully forward model our predictions to this plane, and this quantity is relatively insensitive to obscuration. We show predictions made with our fiducial configuration with $\sigma_{BH} = 0.30$ and the adjusted configuration with $\sigma_{BH} = 0.50$, and for each we show both cases for all predicted AGN and for Compton-thin, *unobscured* AGN, using the L_X-dependent obscuration fraction reported by Ueda et al. (2014). We show results for all predicted AGN and for only Compton-thin AGN, where we adopted the Compton-thick fraction reported by Ueda et al. (2014). In this exercise, we also gain insight into how the AGN population across a wide luminosity range is affected by distinct physical processes. For instance, the number density of the bright AGN is more sensitive to the scatter in the M_{BH}–M_{bulge} relation, as shown in Somerville (2009), because of Eddington bias. Conversely, the fainter populations are more affected by obscuration, as the fraction of X-ray absorbed AGN generally decreases towards higher intrinsic hard X-ray luminosity (Ueda et al. 2003, 2014; Buchner et al. 2015). In addition, the faint end slope is affected by the decay timescale and slope of the AGN lightcurves.

Our predictions are compared to a large compilation of hard X-ray observations (e.g. Ueda et al. 2003, 2014; Aird et al. 2015a,b; Miyaji et al. 2015; Khorunzhev et al. 2018) compiled by Shen et al. (2020, hereafter S20). It is very encouraging that the combined AGN sources and spectral models are able to reproduce the observed evolution of high-redshift AGN in the hard X-ray up to $z \lesssim 6$. However, it is known that it is challenging to produce the observed number density of very luminous AGN at very high redshift ($z \gtrsim 6$) without assuming either more massive seeds or super-Eddington accretion. It is therefore not surprising that our models also show this discrepancy.
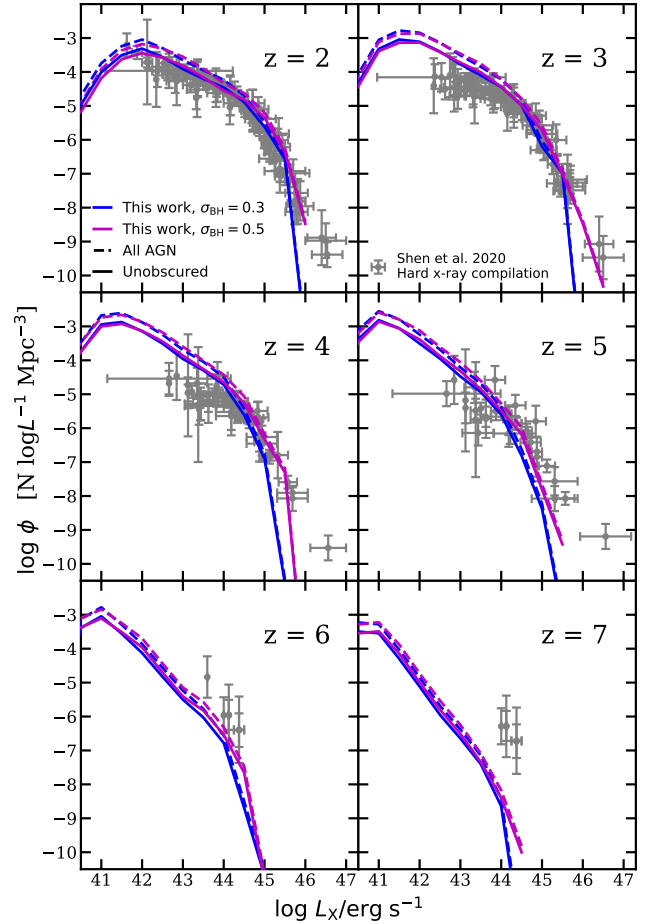
Figure 6. Predicted AGN HXLFs and their evolution with redshift. The lines show the results from our fiducial model ($\sigma_{BH} = 0.3$; blue) and adjusted model ($\sigma_{BH} = 0.3$; purple). The dashed line shows the full predicted AGN population, while the solid line shows only the unobscured AGN population. The data points show the collection of hard X-ray observations compiled by S20.

3.1.2 UV luminosity function

Similarly, Fig. 7 shows the predicted AGN rest-frame UV luminosity functions (UVLFs) between $z = 2$ to 7, where the luminosity is computed by integrating the SED with a tophat filter of width of 200Å centred at 1450Å with inclination dependence included. See bottom panel in Fig. 3 and associated text for a description. Similar to the HXLFs, we show both our fiducial and adjusted models, and we show all AGN and only the Compton-thin ones. For each modelled AGN, an obscured fraction is estimated based on its hard X-ray luminosity and is deducted from the predicted number density per Mpc³.

To forward model our predictions for a direct comparison with observations, we estimate the effect of nuclear scale UV attenuation using the radiation-lifted torus model proposed by Buchner & Bauer (2017). This model provides an estimate (drawn stochastically from a distribution) of the hydrogen column density N_H, based on the AGN X-ray luminosity L_X and M_{BH}, and statistically reproduces the observed fraction of Compton-thick and Compton-thin AGN. We then assume a mean ratio of total neutral hydrogen to
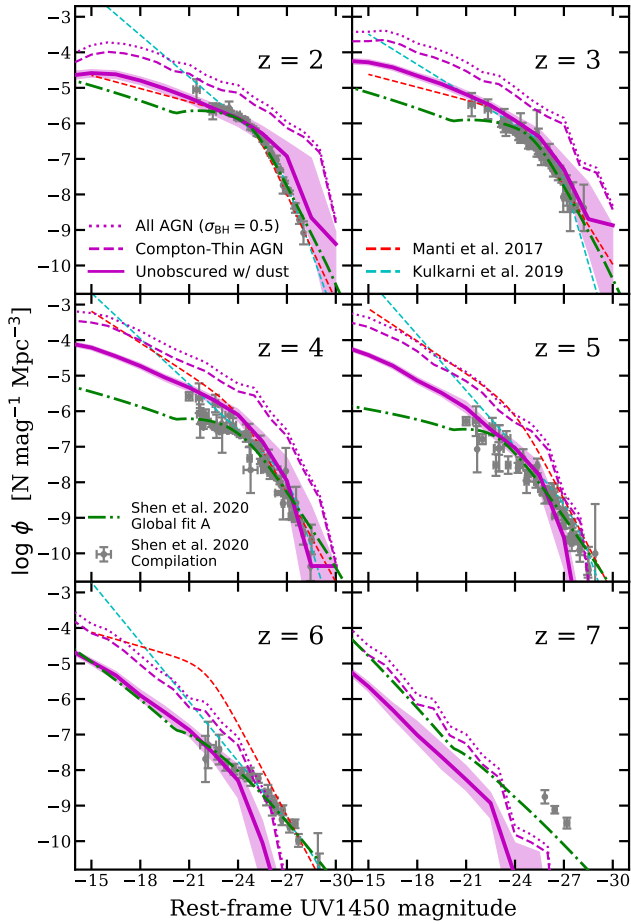
Figure 7. Predicted AGN UVLFs and their evolution with redshift. The lines show the results from our fiducial model ($\sigma_{\rm BH} = 0.3$; blue) and adjusted model ($\sigma_{\rm BH} = 0.3$; purple). The dashed line shows the full predicted AGN population, while the solid line shows only the unobscured AGN populations. The data points show a collection of rest-frame UV1450 observations compiled by S20. The green dot-dashed line shows the Global fit A of S20. We also show fits to observed AGN UVLFs reported by Manti et al. (2017, red dashed line) and Kulkarni et al. (2019, cyan dashed line).

colour excess $N_{\rm H}/E(B-V) = 5.8 \times 10^{21} {\rm cm}^{-2} {\rm mag}^{-1}$ and a total V-band attenuation $R_V \equiv A_V/E(B-V) \sim 5$ for AGN to convert the predicted $N_{\rm H}$ to V-band attenuation A_V (Bohlin et al. 1978; Draine 2003; Gaskell & Benker 2007). This is then converted to attenuation at 1450Å assuming an AGN attenuation curve from Gaskell & Benker (2007). The effect of obscuration and the associated uncertainties are further discussed in Appendix A. Our predicted attenuation correction leads to a considerable decrease in the predicted number density of observable AGN, of an order of magnitude or more. Because the sampled distribution of $N_{\rm H}$ is quite broad (see Buchner & Bauer (2017) and references therein), this leads to a large variability in the attenuation correction and hence on the predicted number density of AGN, especially at the bright end. As there are only a small number of very luminous AGN (both in our modeled sample and in the observed Universe), drawing from this population will not fully sample the distribution of $N_{\rm H}$. In order to obtain

more robust estimates of the attenuation corrected UVLF, we repeat the random draws of $N_{\rm H}$ over 500 independent realizations of our modeled AGN sample, and report the median and the 16th and 84th percentiles in Fig. 7.

These predictions are compared to a large compilation of observational estimates of the AGN UV LF (e.g. McGreer et al. 2013, 2018; Ross et al. 2013; Akiyama et al. 2018; Yang et al. 2018; Wang et al. 2019), compiled by S20. In addition, we show fits to observational results from Manti et al. (2017) and Kulkarni et al. (2019), as well as the 'global fit A' model from S20 for comparison. We note that the faint end of the UVLFs remains largely unconstrained, especially at high redshift. We regard the agreement of our predictions with the available observations as encouraging, although we note that the uncertainties in the obscuration corrections are very large, and could easily result in an order of magnitude uncertainty in AGN number density. However, here we have only included the attenuation correction from the torus scale, neglecting the contribution from the host galaxy scale, which is also expected to be significant (Buchner & Bauer 2017, and references therein). We also show a comparison of these predicted AGN UVLFs to star-forming galaxies UVLFs between $z = 4$ to 7 in Fig. B2.

3.1.3 Bolometric luminosity function

The AGN bolometric luminosity function is a useful space in which to compare different theoretical predictions, and is a useful way to consolidate constraints from many different wavelengths. However, it requires more assumptions to get from the observed quantities to the bolometric luminosity function. For all predicted AGN, we compute the bolometric luminosity, $L_{\rm bol}$, by integrating over the full spectrum for each individual object. In Fig. 8, we show the predicted AGN bolometric LFs at $z = 2$ to 7, made with our fiducial configuration ($\sigma_{\rm BH} = 0.30$) and the adjusted configuration ($\sigma_{\rm BH} = 0.50$). These predictions are compared to a large compilation of constraints derived from observations across a variety of wavelengths. In addition to the aforementioned hard X-ray and UV constraints, this comparison also includes optical (e.g. Shen & Kelly 2012; Palanque-Delabrouille et al. 2013, 2016), mid-IR (e.g. Assef et al. 2011; Singal et al. 2016), and soft X-ray (e.g. Hasinger et al. 2005; Ebrero et al. 2009). S20 converted these multi-wavelength observations to bolometric luminosity by constructing a mean SED template based on observed IR and X-ray SEDs (Hopkins et al. 2007a; Krawczyk et al. 2013) and UV spectral slopes (Lusso et al. 2015). Note that the UV/optical segment of the constructed SED from Krawczyk et al. (2013) is not strongly affected by dust reddening nor obscuration. A correction for obscuration effects from gas and dust is applied (Pei 1992; Morrison & McCammon 1983; Ueda et al. 2014). We show the best-fitting AGN bolometric LF reported by S20 that is fitted to all available constraints and does not restrict the evolutionary pattern for the faint-end slope (labelled as 'global fit A').

Keeping in mind the significant uncertainties involved in the conversions to bolometric luminosity (which are not reflected in the error bars), we find that our model predictions are very consistent with the available observational constraints. It is encouraging that our fiducial model, which is only calibrated to a subset of observations at $z \sim 0$ and
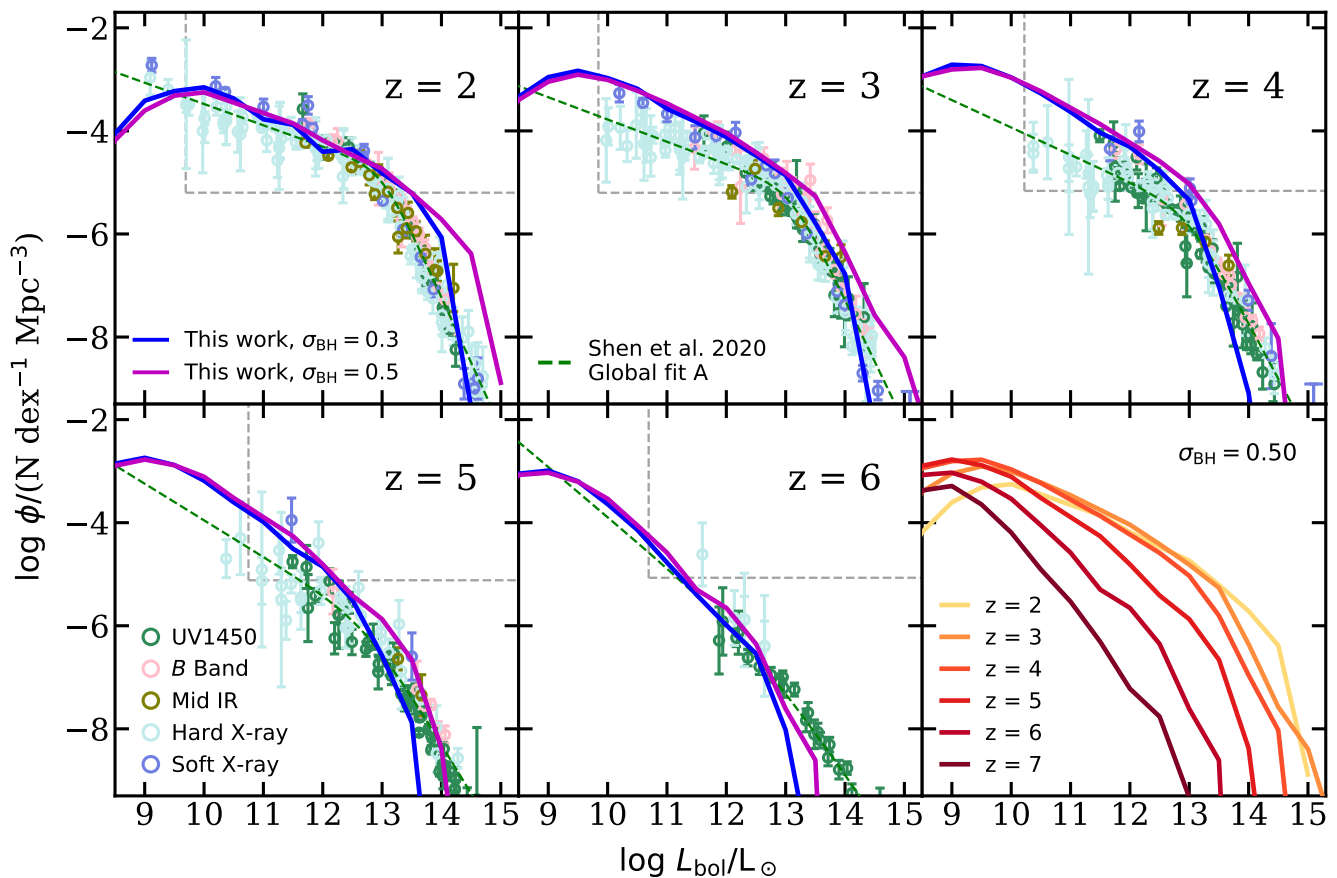
Figure 8. Predicted bolometric AGN luminosity functions and their evolution with redshift between $z = 2$ to 7 for our fiducial model ($\sigma_{\mathrm{BH}} = 0.3$; blue line) and adjusted model ($\sigma_{\mathrm{BH}} = 0.5$; purple line). These predictions are compared to multi-wavelength observational constraints compiled by Shen et al. (2020), which include bolometric corrections and dust and extinction corrections. We also show the fits to the bolometric LFs from S20 (global fit A; green dashed line). The grey dashed lines in each panel indicate the approximate range of objects that are expected to be detected in a typical *JWST* wide-field NIRCam survey, where the vertical boundary loosely corresponds to the detection limit of $m_{\mathrm{F200W,lim}} = 28.60$ and the horizontal line approximately represents where we expect to detect one object for a ~ 100 arcmin^2 survey area in a redshift slice of $\Delta z = 0.5$. The last panel summarizes the predicted evolution of the bolometric LFs in the adjusted model. See text for full descriptions of individual elements included in this plot.

was shown to reproduce a wide variety of observed constraints on galaxy properties from the CANDELS survey (Somerville et al. 2021), is also able to reproduce the observed AGN LFs across a wide luminosity range and their evolution across a wide range of redshift. It also shows that the scatter in the M_{BH}–M_{bulge} relation has a significant effect on the bright end of the AGN LF, as will any other process that introduces scatter into the AGN luminosity. As noted before, the failure of the models to reproduce the brightest AGN at very high redshift is unsurprising, as the formation of these very rare, massive objects may involve physical processes that are not currently included in this model. See Section 5 for a full discussion. The binned luminosity functions presented in this work are available in Table D1 in Appendix D.

Similar to the illustrations presented in Paper I and Paper II, we mark the volume limit and magnitude limit for which objects are expected to be detected in upcoming *JWST* wide-field surveys. The volume limits, marked by the horizontal lines, correspond to the a number density where

one object is expected to be found in a typical wide-field survey with area of ~ 100 arcmin^2, with a redshift slice of $\Delta z = 0.5$ centred at the output redshift. This provides a very rough estimate for when objects become too rare to be expected to be found by *JWST*. However, AGN tend to form in more clustered regions, which is not fully represented in our *volume-averaged* approach to estimating object number density. On the other hand, the vertical lines indicate an approximate detection limit of $m_{\mathrm{F200W,lim}} = 28.6$, calculated assuming an exposure time comparable to past *HST* wide-field surveys. This observed-frame IR magnitude limit is converted to rest-frame UV at each output redshift for the predicted galaxy populations based on the results presented in Paper I. This rest-UV limit is then translated to L_{bol} based on a subset of predicted AGN selected within a narrow bin around the limiting M_{UV}. We note that identifying high-redshift AGN requires selection methods very different from the ones for galaxies (Volonteri et al. 2017). This exercise simply provides a rough estimate for which
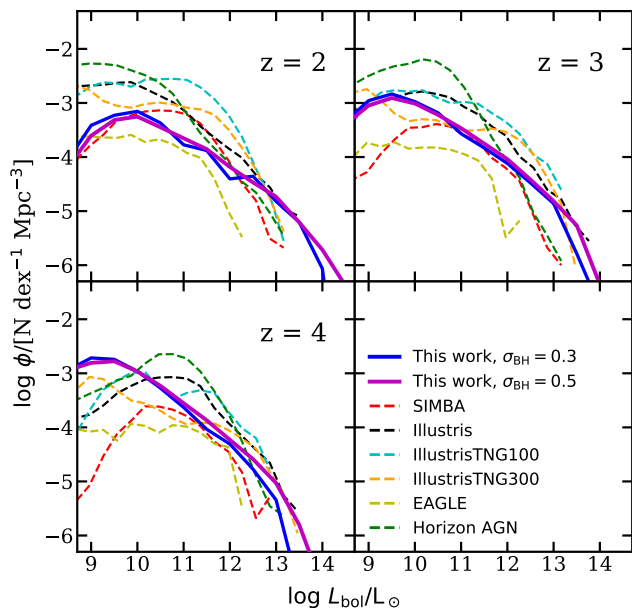
Figure 9. The predicted AGN bolometric luminosity functions from our fiducial model ($\sigma_{\mathrm{BH}} = 0.3$; blue) and adjusted model ($\sigma_{\mathrm{BH}} = 0.5$; purple) at $z = 2$, 3, and 4, compared to a compilation of modern cosmological hydrodynamic simulations, including SIMBA (Davé et al. 2019), TNG–100 anf TNG–300 (Nelson et al. 2019), EAGLE (Schaye et al. 2015), Horizon AGN (Dubois et al. 2014), and Illustris (Genel et al. 2014). See text for details.

AGN populations are above the detection limit in *JWST* galaxy surveys.

In Fig. 9, we compare the same bolometric LFs to the results from a selection of state-of-the-art cosmological hydrodynamic simulations, including SIMBA (Davé et al. 2019), Illustris (Genel et al. 2014), IllustrisTNG–100 and IllustrisTNG–300 (Nelson et al. 2019), the Evolution and Assembly of GaLaxies and their Environments (EAGLE) simulations (Schaye et al. 2015), and Horizon–AGN (Dubois et al. 2014). The results were compiled by Habouzit et al. (in preparation) and we refer the reader to that work and Habouzit et al. (2021) for in-depth discussion of the simulations and a comparison with observations. These simulations have comparable simulated volume and mass resolution, and are optimised to simulate the formation of galaxies and AGN in a cosmological volume. For instance, SIMBA and Horizon AGN are 100 h^{-1}Mpc on a side, with particle mass of the order of $\sim 10^8$ M$_\odot$. Illustris, IllustrisTNG–100, and EAGLE are ~ 100 Mpc on a side, with particle mass of $\sim 10^7$ M$_\odot$. IllustrisTNG–300 has the largest simulated volume among all compared simulations of 302.6 Mpc on a side, with a relatively coarse dark matter particle mass of 5.9×10^7M$_\odot$. These simulations each adopt very different prescriptions and parametrizations for baryonic physics, and have adopted different calibration strategies, such as which observational constraints and redshifts to prioritize for the calibration, as well as what redshift range to cover. For example, some simulations may terminate at high redshifts and are not calibrated to observational constraints from the local universe. The bolometric luminosities for AGN in these numerical simulations are calculated based on the predicted

black hole accretion rate assuming the standard relation $L_{\mathrm{bol}} = \epsilon_r \dot{M}_{\mathrm{acc}} c^2$, where ϵ_r is the radiative efficiency. The radiative efficiency is set to $\epsilon_r = 0.25$ for Illustris and TNG–100; and $\epsilon_r = 0.11$ for Horizon-AGN, EAGLE, and SIMBA. These parameters are chosen to be consistent with the modelling choices made in the simulations. See Habouzit et al. (2021) and Habouzit et al. (in preparation) for a detailed discussion. We note that some studies adopt an alternative parametrization that replaces ϵ_r with $\epsilon_r/(1 - \epsilon_r)$. We also compute ϵ_r based on predictions from our physically-grounded modelling pipeline and compare to values assumed in previous studies in Appendix C. From this comparison, we find that our predictions are in broad agreement with numerical simulations, with some simulations having steeper faint end slopes and higher AGN number densities, and others having shallower faint end slopes and lower AGN number densities than our models predict. We note that the numerical simulations probably underestimate the number density of low-luminosity AGN, particularly at high redshift, due to their limited mass resolution, and may have large uncertainties in the number density of luminous AGN, due to their limited volume. Overall, it is encouraging that our SAM-based predictions are consistent with those from numerical hydrodynamic simulations.

3.2 Hard ionizing photon emissivity and helium reionization history

The reionization of intergalactic helium is a cosmological-scale phase transition driven mainly by AGN. Currently, huge uncertainties remain in the estimates of the helium ionizing photon budget, which determines the subsequent progression of the cosmic helium reionization history. In this subsection, we present estimates for the total helium ionizing photon budget available during the EoR, based on the predicted AGN number density and spectral characteristics from our physics based models.

In Fig. 10, we show the comoving specific 1 ryd (non-ionizing) photon budget, ϵ_{912}, and the comoving He II ionizing photon budget ($\lambda < 228$Å), $\dot{n}_{\mathrm{ion,HeII}}$, as a function of redshift. Here $\dot{n}_{\mathrm{ion,HeII}}$ is calculated with equation (4) assuming an escape fraction of unity and no obscuration, to account for all ionizing photons produced by AGN. These estimates are essentially a combination of the results from the AGN populations from Fig. 8 and the individually predicted emissivity shown in Fig. 4 and Fig. 5. We show results from both the fiducial ($\sigma_{\mathrm{BH}} = 0.3$) and adjusted ($\sigma_{\mathrm{BH}} = 0.5$ model, and find that the additional population of luminous AGN in the adjusted model can contribute nearly a factor of two more ionizing photons towards lower redshifts. The tabulated ionizing photon budget is provided in Table 1. We also note that ϵ_{912} may not be a very good tracer for $\dot{n}_{\mathrm{ion,HeII}}$, given that high-redshift AGN tend to accrete at a higher accretion rate than their low-redshift counterparts with comparable black hole masses, which may have an impact on the overall spectral slope and introduce an effective redshift-evolution in the ratio between UV and bolometric luminosities.

In Fig. 11, we show the distribution between ϵ_{912} and $\dot{n}_{\mathrm{ion,HeII}}$ and between ϵ_{912} and the AGN He II ionizing production efficiency, defined as $\xi_{\mathrm{ion,AGN}} \equiv \dot{N}_{\mathrm{HeII}}/\epsilon_{912}$, for the predicted AGN populations at $z = 3$. The physical quanti-
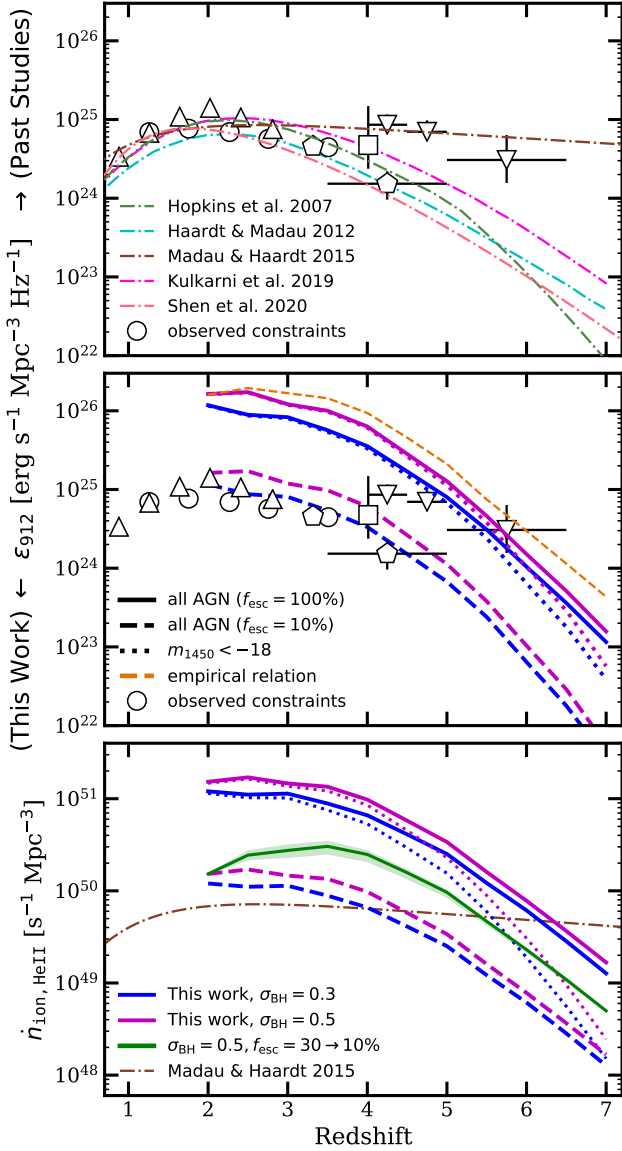
Figure 10. *Top and middle panels:* Redshift evolution of the cosmic specific emissivity at 1 ryd, ϵ_{912}, between $z = 2$ to 7 predicted by our fiducial model ($\sigma_{BH} = 0.3$; blue) and adjusted model ($\sigma_{BH} = 0.5$; purple). The solid line shows all AGN and the dashed line only includes objects with $M_{UV} \leq -18$, similar to the other studies shown in this comparison. The orange dashed line labelled 'empirical relations' shows the emissivity calculated using the Hassan et al. (2018) scaling relation as shown in Fig. 4. Our results are compared to past analytic studies by Madau & Haardt (2015) and estimates from fits to observed AGN UVLFs (Hopkins et al. 2007a; Haardt & Madau 2012; Shen et al. 2020). We also show a compilation of observations, including Bongiorno et al. (2007, circles), Glikman et al. (2011, squares), Masters et al. (2012, pentagons), Palanque-Delabrouille et al. (2013, triangles), Giallongo et al. (2015, upside-down triangles). *Bottom panel:* He II ionizing emissivity, $\dot{n}_{ion,He\,II}$, calculated by integrating individual AGN SEDs at > 4 ryd assuming an escape fraction of unity. We show additional cases with constant $f_{esc}=0.10$ (purple and blue dashed lines), and with f_{esc} evolving as a function of redshift 0.30 at $z \gtrsim 7$ to 0.10 at $z \sim 2$ (green line). The shaded regions correspond to a range of evolutionary paths for the escape fraction. See text for details.
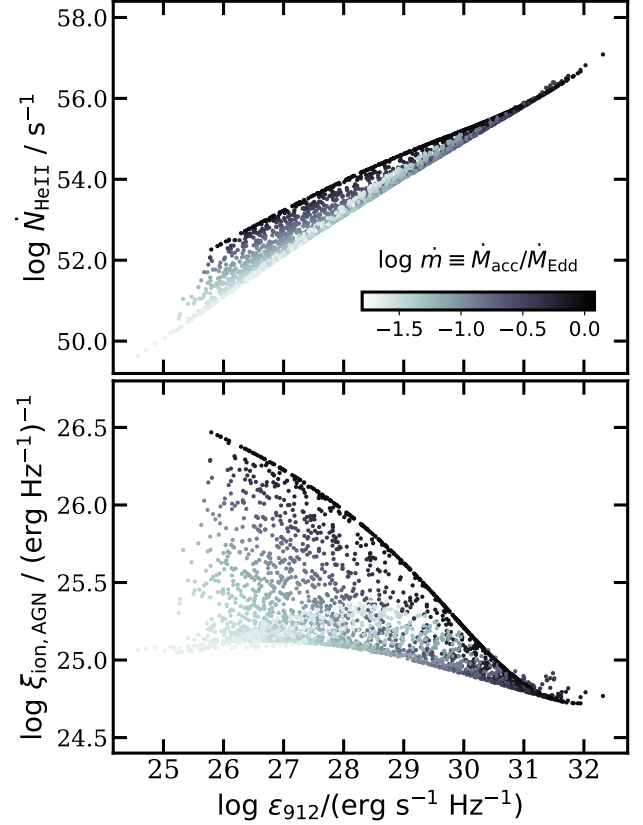

Figure 11. Distribution of ϵ_{912} versus $\dot{N}_{ion,He\,II}$ (*top panel*) and $\xi_{ion,AGN}$ (*bottom panel*) for our simulated AGN population, colour-coded by the Eddington-normalized accretion rate. Here we show AGN from at output at $z = 3$; however, this relation does not explicitly depend on redshift.

ties shown in this plot and the model components related to these predictions do not do not have any explicit redshift dependence. For both relations, the top and bottom boundaries are marked by the Eddington accretion rate and the cut-off accretion rate where AGN become radiatively inefficient. Both panels show that the scatter in these relations arises mainly from the BH accretion rate, which has a strong effect on the shape of the AGN spectra (see top panel in Fig. 3). We also predict that this scatter increases rapidly towards the UV-faint AGN populations, where the scatter in $\xi_{ion,AGN}$ can be over a dex for AGN with $\epsilon_{912} \lesssim 28$. Therefore, ϵ_{912} may not be an effective tracer for the hard ionizing photon production rate. Fig. 12 is a similar figure where we show the conditional distributions of $\xi_{ion,AGN}$ versus ϵ_{912}. These 2D histograms are colour-coded to reflect the number density weighted distribution in each of the corresponding (vertical) ϵ_{912} bins among AGN populations. Furthermore, while Fig. 11 shows that there is a large scatter in the ϵ_{912}–ξ_{ion} relation, the number density weighted distributions shown in Fig. 12 break down the contribution from AGN across a range of redshifts. When accounting for the evolution between AGN number densities and their accretion rate (similar to that of Fig. 2), we see that ϵ_{912} is even less reliable as a tracer for the collective contribution from all AGN (e.g. cosmic ionizing emissivity $\dot{n}_{ion,He\,II}$) given

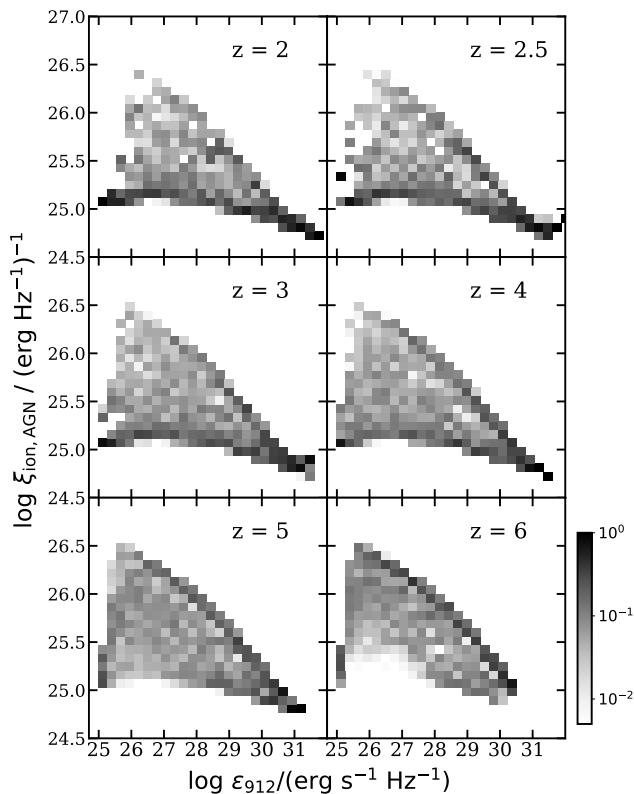**Figure 12.** Conditional distributions of $\xi_{\mathrm{ion,AGN}}$ versus ϵ_{912} between $z = 2$ to 6, predicted by our fiducial model. The 2D histogram is colour-coded with the conditional number density per Mpc^3 of AGN in each bin, normalized to the sum of the number density in the corresponding (vertical) ϵ_{912} bin.

the redshift evolution in BH accretion rates across the AGN populations.

In our estimates of these *photon budgets*, we adopt the photon production rate contributed by all predicted physical sources, without accounting for the effects from dust and gas obscuration. Instead, these effects will be folded into the overall ionizing photon escape fraction. This is a modelling decision we made with the intent to keep a clean separation between the *production* of ionizing photons by the full population of sources, the *fraction* of ionizing photons that were able to escape to the IGM, and whether or not a given AGN is *detected* at a specific wavelength with a specific detection limit. We emphasize that this is different from the conventional *emissivities* reported by other studies, which only account for the *detected* AGN populations at their observed luminosities. For instance, Hopkins et al. (2007a), Haardt & Madau (2012), Kulkarni et al. (2019), and S20, estimate the 1450Å emissivity (ϵ_{1450}) by integrating observationally fitted AGN UVLFs over a certain luminosity range (see Fig. 7), where extinction and obscuration are implicitly accounted for. This is then converted to 912Å emissivity assuming a power-law spectrum. On the other hand, Madau & Haardt (2015) provides an empirical estimate that is motivated by the set of observational constraints that are shown in Fig. 10, which are also derived from a set of FUV (Bongiorno et al. 2007; Glikman et al. 2011; Masters et al. 2012; Giallongo et al. 2015) and optical observations (Bongiorno

et al. 2007; Palanque-Delabrouille et al. 2013), where similar extrapolation and power-law assumptions were applied. In effect, *these previous studies assume that AGN that are not detected at 1450Å do not contribute anything to the ionizing photon budget.* Yet, we have shown that torus-scale obscuration, which is expected to be highly anisotropic, can easily decrease the normalization of the observed 1450Å LF by an order of magnitude or more. Although an AGN with a torus viewed at a small inclination angle may be so obscured along our line of sight that it drops out of a UV-selected survey, it is expected that photons, including ionizing photons, should be able to escape along directions within some solid angle perpendicular to the torus, thus potentially contributing to the global budget of ionizing photons. Indeed, there is strong observational evidence that ionizing radiation from quasars can be highly anisotropic (e.g. Lau et al. 2017). This difference largely explains why our predictions are approximately an order of magnitude higher than previous predictions from the literature, which were all based on observed UV 1450Å luminosity functions. In addition, it can be seen from Fig. 7 and Fig. 8 that the predictions of our physics-based models tend to yield steeper faint end slopes for the AGN LFs than the assumed values that are extrapolated in previous studies. Finally, as we have shown, our physically based AGN spectral synthesis model yields somewhat different results than the commonly assumed fixed power law spectrum or template SED. To illustrate the impact of adopting conventional conversions assuming power-law spectra and simplified scaling relations, we show in Fig. 10, an alternative comoving emissivity ϵ_{912} calculated assuming a scaling relation between M_{BH} to B-band luminosity and a power-law conversion to ϵ_{912} similar to the one used in Hassan et al. (2018, also see the orange dashed line in Fig. 4). We find that this yields an overestimate in the comoving emissivity of about a factor of two relative to our full physics-based model predictions.

To obtain the predicted ionizing photon budget, we apply an overall escape fraction that folds in the effects of attenuation and obscuration, which is applied to all AGN when modelling the comoving hard ionizing photon emissivity and the subsequent volume filling fraction of He III. As reported by numerical simulations, the escape fraction for star-forming galaxies can be very stochastic depending on the internal structures and the many intricate physical processes occurring in individual halos and even individual star forming regions (Paardekooper et al. 2015; Xu et al. 2016; Ma et al. 2020) and some similarities can be drawn for active galaxies. We add that the escape fraction of He ionizing photon from AGN is a highly complex, unsolved problem that has been greatly understudied. Given the large uncertainties and stochasticity of this quantity both across AGN populations and across time, we adopt a simple empirical approach similar to the one adopted in Paper IV and treat the escape fraction as a population averaged quantity, which is the escape fraction of all hard ionizing photons collectively produced by all AGN. In this work, we allow f_{esc} to be a constant value or to vary as a function of redshift. We are in the process of developing a more advanced escape fraction model that accounts for the physical properties of individual sources (Yung et al. in preparation).

In this work, we consider two bracketing cases where we allow a maximum escape fraction of $f_{\mathrm{esc}} = 100\%$ and a mini-
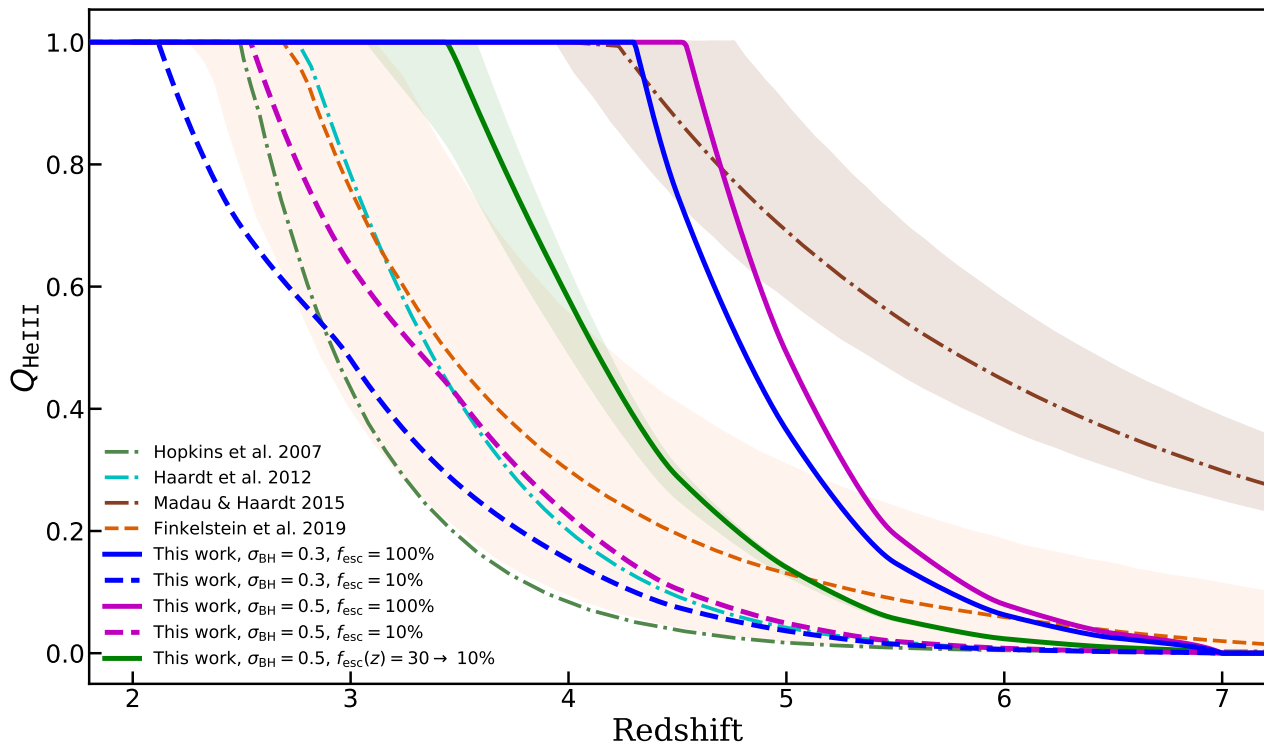
Figure 13. The He III volume filling fraction, $Q_{\mathrm{He\,III}}$, as a function of redshift. Solid blue and purple lines show predictions from our fiducial model ($\sigma_{\mathrm{BH}} = 0.3$) and adjusted model ($\sigma_{\mathrm{BH}} = 0.5$) with an escape fraction of unity, and blue and purple dashed lines show the predictions with a constant $f_{\mathrm{esc}} = 0.1$. We also show two additional cases for the adjusted model, where f_{esc} evolves as a function of redshift from 0.30 at $z \gtrsim 7$ to 0.10 at $z \sim 2$ (green). The shaded regions correspond to a range of evolutionary paths for the escape fraction. See text and Fig. 14 for details. These cases are colour-matched to the escape fraction evolution plotted in Fig. 14. Our results are compared with past studies by Hopkins et al. (2007a), Haardt & Madau (2012), Madau & Haardt (2015), and Finkelstein et al. (2019), as shown in the figure label.
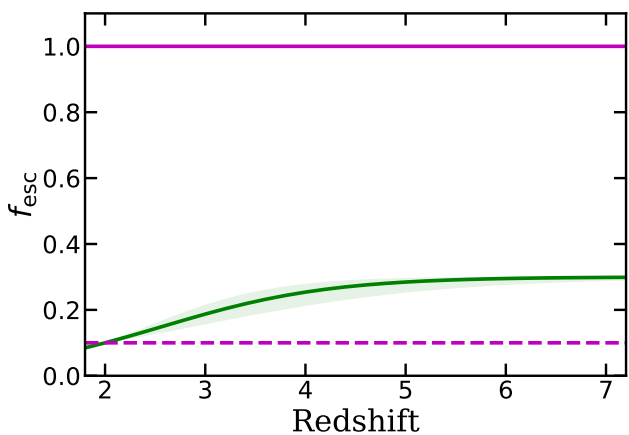

Figure 14. Illustration of how escape fraction evolves with redshift in the various scenarios we consider in this work. The green line shows a scenario where the escape fraction evolves from 0.30 to 0.10 between $z \sim 7$ and 2 according to Eqn. 8 with $k = 1.2$. The top and bottom boundaries of the green shaded region show alternative scenarios with $k = 1.6$ and $k = 0.8$ to illustrate how sensitive the model output is to the assumed escape fraction. The purple solid and dashed lines show a constant escape fraction of $f_{\mathrm{esc}} = 1.00$ and 0.10.

mum escape fraction of $f_{\mathrm{esc}} = 10\%$. In addition, we consider a redshift-dependent average escape fraction modelled with the logistic equation as presented in Paper IV:

$$f_{\mathrm{esc}}(z) = \frac{f_{\mathrm{esc,max}}}{1 + \left(\frac{f_{\mathrm{esc,max}}}{f_{\mathrm{esc,0}}} - 1 \right) e^{-k_0(z - z_0)}}, \qquad (8)$$

assuming f_{esc} decreases from $f_{\mathrm{esc,max}}$ at high redshift at a characteristic rate k_0 until it asymptotically reaches $f_{\mathrm{esc,min}}$ at an anchoring redshift z_0. This can be loosely interpreted as a *population-averaged* quantity, where all AGN share the same value, or as the effective escape fraction of the total number of ionizing photon collectively produced by all AGN. This approach allows us to roughly estimate the plausible range for the global escape fraction in order to reproduce observational constraints.

Based on this parametrization, we explore a scenario which the *population-averaged* escape fraction evolves from a minimum value $f_{\mathrm{esc,min}} = 0.1$ at $z_0 = 2$ to $f_{\mathrm{esc,max}} = 0.30$ at $k_0 = 1.2 \pm 0.4$. This is illustrated in Fig. 14, along with the two other non-evolving cases of $f_{\mathrm{esc}} = 1.0$ and 0.1 considered in this section. Here, $f_{\mathrm{esc,max}} = 0.30$ is chosen very loosely to represent a more realistic approximation of ionizing photons are escaping averaged across all AGN. We note that the value of $f_{\mathrm{esc,max}}$ adopted here is in rough agreement with the fraction of unobscured AGN at $z \sim 2$ reported in Buchner et al. (2015). The growth rate $k_0 = 1.2$ is empirically set by hand to reproduce a relatively smooth transition

Table 1. Tabulated data for the He II ionizing emissivity $\dot{n}_{\mathrm{ion,He\,II}}$ (s^{-1} Mpc^{-3}), specific emissivity ϵ_{912} (erg s^{-1} Mpc^{-3} Hz^{-1}), and reionization history between $z = 2$ to 7, assuming an ionizing photon escape fraction of 100% and 10%.

| | | | $Q_{\mathrm{He\,III}}$ $\sigma_{\mathrm{BH}} = 0.5$ | |
redshift	$\log \epsilon_{912}$	$\log \dot{n}_{\mathrm{ion,He\,II}}$	f_{esc}=100%	f_{esc}=10%
2.0	26.21	51.18	1.00	1.00
2.5	26.24	51.23	1.00	1.00
3.0	26.08	51.17	1.00	0.63
3.5	26.00	51.13	1.00	0.42
4.0	25.80	50.99	1.00	0.23
4.5	25.45	50.76	1.00	0.11
5.0	25.10	50.53	0.49	0.05
5.5	24.66	50.20	0.19	0.02
6.0	24.18	49.89	0.08	0.01
6.5	23.71	49.56	0.03	0.00
7.0	23.20	49.22	0.00	0.00

from $f_{\mathrm{esc,min}}$ to $f_{\mathrm{esc,max}}$. The upper bound of the shaded regions in Fig. 14 corresponds to $k_0 = 1.6$, where f_{esc} remain at $f_{\mathrm{esc,max}}$ longer before dropping down to $f_{\mathrm{esc,min}}$. Conversely, the lower bound corresponds to $k_0 = 0.8$, which f_{esc} at a more rapid pace. We also note that while the parametrization for AGN f_{esc} is the same as the one we adopt for star-forming galaxies in Paper IV, the parameters for the escape fraction for AGN populations are set independent of our previous work.

In Fig. 13, we show predictions for the redshift evolution of the *volume-averaged* IGM He III volume filling fraction, $Q_{\mathrm{He\,III}}$ obtained by solving equation (6). We show the two bracketing cases of $f_{\mathrm{esc,He}} = 100\%$ and 10%. While the former is unlikely given that a large fraction of galaxies are known to be obscured, this gives an estimate of the absolute upper limit of how quickly helium could have been reionized given the total amount of ionizing photons produced. On the other hand, adopting f_{esc}= 10% yields results that are generally in good agreement with estimates reported by past studies, and are consistent with observational constraints indicating that He II reionization is likely still in progress at $z \sim 2.7$ (McQuinn et al. 2009; Shull et al. 2010; Syphers & Shull 2014). In addition, H I Lyα forest measurements indicate a bump in the IGM thermal history at $z \sim 2.8$, which has been interpreted as an indirect indicator for the end of He II reionization (Schaye et al. 2000; Becker et al. 2011; Becker & Bolton 2013; Puchwein et al. 2015; Upton Sanderbeck et al. 2016; Hiss et al. 2018). We note that these broad indicators do not constrain the early stages of the phase transition nor the variance due to the overall patchiness of the process. We find that either of our redshift-dependent f_{esc} scenarios are capable of reionizing the IGM well before $z \sim 4$. The tabulated data for the predicted ionizing emissivity and reionization history are summarized in Table 1.

We also compare our predictions with past studies that adopt a similar reionization model. The Finkelstein et al. (2019) model allows a large degree of freedom in the AGN comoving 1-ryd specific emissivity spanning a range between the UVLF-based result from Hopkins et al. (2007a) and the AGN-dominated result from Madau & Haardt (2015) (see Fig. 10). The AGN emissivity and subsequent reionization history, along with the high-redshift galaxy UVLFs and
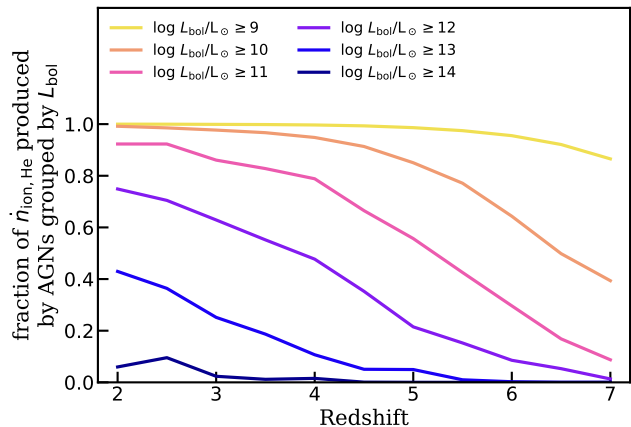

Figure 15. Predicted redshift evolution for the cumulative fraction of He II ionizing photons contributed by AGN above various bolometric luminosity limits. These predictions show the intrinsic production rate and do not account for the escape fraction that possibly varies across the AGN population. These estimates can be used to assess the completeness of ionizing photon estimates in surveys where faint AGN are not fully sampled.

the Lyman-continuum production efficiency, are then jointly constrained by a collection of observational constraints using a Markov Chain Monte Carlo (MCMC) approach. The orange line and shaded region show the median and the 68% central confidence range from their fiducial model for comparison. The line labelled Hopkins et al. 2007a is a special case from the (Finkelstein et al. 2019) model where the AGN 1 ryd emissivity is restricted to follow the results of Hopkins et al. (2007a). We note that our model configurations (IGM mean comoving helium number density and recombination time-scale) are consistent with the assumptions in Finkelstein et al. (2019). The Haardt & Madau (2012) analytic model also adopts an AGN ϵ_{912} that is in close agreement with the results of (Hopkins et al. 2007a), which provides a good example of how even with the same assumed underlying AGN population, modelling assumptions adopted by different studies, such as the conversion to $\dot{n}_{\mathrm{ion,He\,II}}$ and escape fraction, may impact the final predicted reionization history. The Madau & Haardt (2015) model presents an extreme scenario where AGN completely reionize both intergalactic hydrogen and helium without any contribution from galaxies. We show results from their default model configurations, as well as the range corresponding to changes in clumping factor, IGM temperature, and EUV spectral slope. We find that even our most 'optimistic' model with $f_{\mathrm{esc,\,He}} = 1$ does not produce the very early helium reionization predicted by the Madau & Haardt (2015) model. Our fiducial and adjusted models with $f_{\mathrm{esc,\,He}} = 0.1$ produce predictions for the helium reionization history that are within the range of previous studies such as Hopkins et al. (2007a), Haardt & Madau (2012), and Finkelstein et al. (2019).

An advantage of our physics based model, which predicts the ionizing photon rate for each galaxy and halo, is that we are able to further examine the contribution of ionizing photons from AGN of different halo masses and bolometric luminosities. In Fig. 15, we break down the contribution from AGN by bolometric luminosity between $z = 2$ to 7.
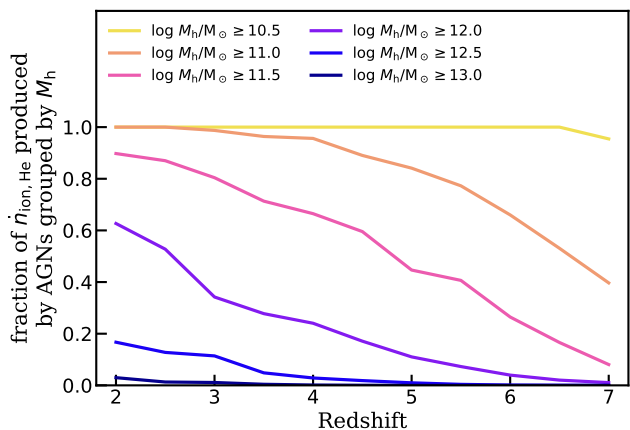
Figure 16. Predicted redshift evolution for the cumulative fraction of He II ionizing photons contributed by AGN hosted in halos above various halo mass limits. These predictions reflect the intrinsic production rate and do not account for the escape fraction, which may depend indirectly on halo mass. This can be used to assess the completeness of ionizing photon estimates in simulations where the full dark matter halo population is not fully sampled due to mass resolution limits or limited simulation volume.

Figure 17. The production rate of He II ionizing photons, $\dot{N}_{\mathrm{ion,He\,II}}$, as a function of AGN host halo mass between $z = 2$ to 6, predicted by our fiducial model. The 2D histogram is colour-coded with the conditional number density per Mpc^3 of AGN in each bin, normalized to the sum of the number density in the corresponding (vertical) halo mass bin. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The last panel shows an overlay of the median $\dot{N}_{\mathrm{ion,He\,II}}$ values at $z = 2$, 4, and 6 from our model.

This intuitively illustrates which AGN are dominating the production of hard ionizing photons at any given time. As expected, at earlier times the majority of ionizing photons are dominated by the more prevalent faint AGN, which then gradually transition to the brighter AGN as they grow in number density. These results can also be used to estimate the completeness of ionizing sources captured by a given survey. For example, luminous AGN above the detection limit for a typical *JWST* wide-field survey at $\log(L_{\mathrm{bol}}/L_\odot) \sim 11$, as shown in Fig. 8, are collectively producing $\sim 40\%$ of the total ionizing photons. Note that these cuts are made based on bolometric luminosity and do not account for the survey volume, which is a separate requirement for properly capturing the very rare high-redshift quasars.

Similarly, Fig. 16 shows a breakdown for the contribution of He ionizing photons as a function of the host halo masses. This is helpful for assessing the fraction of ionizing photons that are captured in numerical simulations, where low-mass halos may fall below the mass resolution.

Furthermore, we investigate how the production rate of hard ionizing photons scales with the AGN host halo mass. In Fig. 17, we present the predicted scaling relation and distribution for He II ionizing photon production rate, $\dot{N}_{\mathrm{ion,He\,II}}$, with respect to the host halo mass between $z = 2$ to 6. Tabulated data for the median and scatter are given in Table D2. We find that the redshift evolution in this relation is driven by a subtle increase in both $\dot{M}_{\mathrm{acc}}$ and M_{BH} in halos of similar mass towards lower redshifts, which both contribute to a higher ionizing photon count for a halo at a given mass. We note that the most massive halos presented in this work are extremely rare, and their contributions to the overall ionizing photon budget are extremely low as illustrated in Fig. 16. We also note that Fig. 17 only includes radiatively efficient AGN ($\log \dot{m} \gtrsim -1.7$) and their host halos. Halos with a radiatively inefficient AGN or no AGN are not included. These results can be adopted in (semi-)numerical simula-
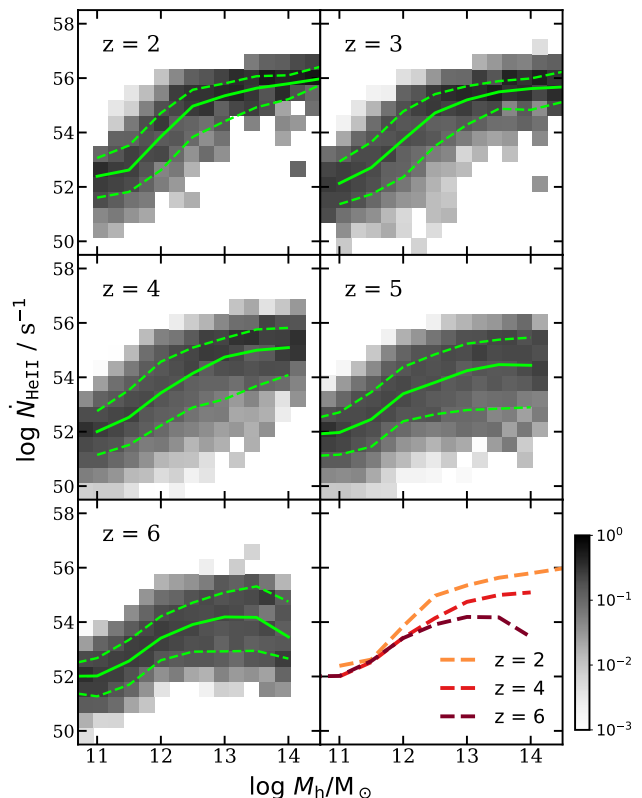
tions that assume simple scaling relations between ionizing photon production rate and stellar mass or halo mass (e.g. Hassan et al. 2016, 2018). This result is analogous to one of the results presented in Paper III, where we explored the ionizing photon production rate by galaxies as a function of halo mass.

4 AGN CONTRIBUTION TO COSMIC HYDROGEN REIONIZATION

In Paper III and Paper IV, we presented predictions from a detailed physical model showing estimates for the production of ionizing photons from high-redshift galaxies and the implied reionization history of intergalactic hydrogen. The results indicate that these galaxies, assuming a Lyman-continuum (LyC) escape fraction that is broadly consistent with other studies, are likely to have produced sufficient ionizing photons to fully reionize the Universe within the time-frame consistent with the observed IGM and CMB constraints. In these studies, the contribution from AGN at $z \gtrsim 4$ were assumed to be negligible. In this section, we provide estimates for the contribution to hydrogen reioniza-

tion from AGN based on the new inclusion of AGN in our modelling infrastructure.

The hydrogen ionizing photon production rate, $\dot{N}_H$, for individual AGN is obtained by integrating equation (4) between 912Å and 228Å, or equivalently from 13.6 to 54.4 eV, and the overall contribution from all AGN is calculated using equation (5), with the helium ionizing photon production rate replaced by $\dot{N}_H$. In order to assess the contribution from AGN at low redshift, we have extended the predictions from previous studies down to $z = 2$ based on the same model configurations and calibrated parameters as presented in Paper I and Paper IV. In Fig. 18, we show the fraction of H ionizing photons produced by AGN relative to the combined contributions from AGN and galaxies as a function of redshift. We also compare this to the contribution from galaxy populations broken down by rest-frame UV luminosity, which can be directly compared to fig. 12 in Paper IV. We see that the contribution from AGN during the epoch of hydrogen reionization, $6 \lesssim z \lesssim 10$, is fairly negligible, but the contribution rapidly rises to ~ 30 per cent at $z \sim 4$ and can even overtake that from all galaxies at $z \lesssim 3$. We note that this plot shows the *intrinsic* production rate of ionizing photons, and does not account for the escape fraction, which may differ between AGN and galaxies, and may vary as a function of halo mass and other galaxy properties.

Furthermore, we show the comoving hydrogen ionizing emissivity from AGN as a function of redshift in Fig. 19, assuming a fixed $f_{esc} = 1.00$ and 0.10 for AGN. To compare this with the contribution from galaxies, we include one of the key results from Paper IV, where a Markov Chain Monte Carlo (MCMC) pipeline was implemented to determine the redshift-evolution of escape fraction for star-forming galaxies, $f^{\star}_{esc}$, to match the observed Lyα constraints from Becker & Bolton (2013) and the Thomson scattering optical depth of the CMB, τ_{CMB}, reported by Planck Collaboration (XLVII 2016b). In this previous study, we reported that a LyC escape fraction that evolves from $f^{\star}_{esc} \sim 25$ per cent at $z \gtrsim 12$ to a few per cent towards the conclusion of reionization at $z \sim 6$ results in good agreement with all available constraints on the reionization history. See section 3.2 in Paper IV for more details.

Although our models predict that it is unlikely that AGN have contributed significantly to the initial reionization of intergalactic hydrogen, we note that their contribution increases more rapidly than that of galaxies and could be comparable to that of galaxies at around $z \sim 4$. If this additional AGN contribution were accounted for in the MCMC framework as in Paper IV, it would likely result in even lower values for the implied escape fraction from galaxies at these redshifts (depending, of course, on the corresponding escape fraction from AGN). With fairly plausible assumptions for the escape fraction, the combined contribution from AGN and galaxies is consistent with the Becker & Bolton constraints at $z \sim 2$-4. The fairly steep rise in emissivity due to the AGN contribution may provide an explanation for the bump at $z \sim 3$. Furthermore, AGN could serve as an 'insurance policy' in case the escape fractions in galaxies are found to be lower than those favored in our models. In this case, the contribution from AGN could help complete hydrogen reionization.

Untangling the degenerate contribution from AGN and galaxies towards the end of cosmic hydrogen reionization has
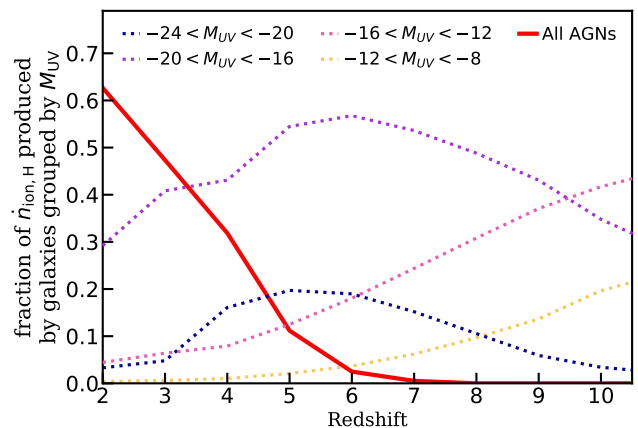

Figure 18. Predicted fraction of cosmic hydrogen ionizing emissivity, $\dot{n}_{ion,H}$, contributed by AGN between $z = 4 - 7$ (red solid line) relative to the combined number of ionizing photons produced by AGN and all star-forming galaxies. Similar to fig. 12 presented in Paper IV, the contributions from star-forming galaxies are divided into groups by their rest-frame dust-attenuated M_{UV} (dotted lines). Galaxies outside the range of M_{UV} shown contribute < 1 per cent of ionizing photons at all times.

proven to be very difficult with the current set of constraints on IGM thermal history, given that each class of objects has their own set of uncertainties in number density, ionizing production rate, and ionizing photon escape fraction. In this work and the rest of the series, we have demonstrated that a detailed physically-based source model that self-consistently models the co-evolution of galaxies and their black holes can provide a novel approach to gaining insights into this puzzle. Future work on more physically grounded modelling of ionizing photon escape fraction together with anticipated constraints on the faint galaxy nad AGN population that will be obtained by *JWST* will bring further insights that will help untangle the contribution from AGN and galaxies.

5   DISCUSSION

This *Semi-analytic forecasts for JWST* paper series aims to provide a comprehensive collection of predictions for observable properties and underlying physical properties of objects forming in the early universe. Predictions from these successful models presented earlier in the series were used to make detailed predictions for the expected properties of galaxies that have been used to facilitate the planning of upcoming surveys, as well as in the development of data reduction and interpretation pipelines. This latest addition to the work series expands our framework to include high-redshift AGN populations and establish the connections between the predicted *intrinsic* AGN properties and their spectral properties, as well as the impact on the reionization history.

The semi-analytic approach represents a 'middle way' between empirical or semi-empirical models and numerical simulations. This approach uses a selection of models and recipes to track the physical processes that occur at vastly different physical scales. This physically motivated approach preserves the physical origins behind the predicted phenomenological relations and enables us to connect these
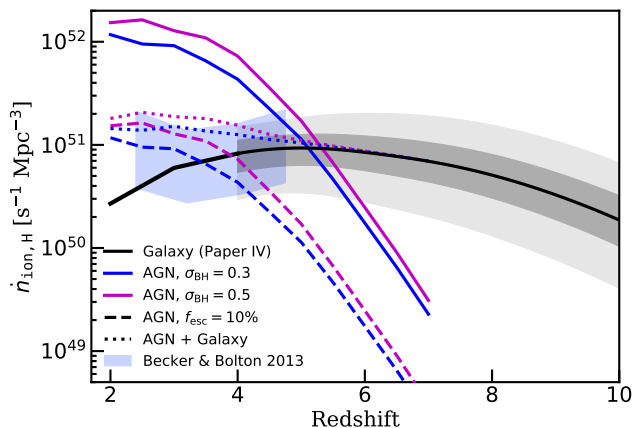
Figure 19. Hydrogen ionizing emissivity contributed by AGN and galaxies as a function of redshift. The purple (blue) line shows predictions from our adjusted (fiducial) model, where the solid and dashed lines assumes AGN escape fraction of f_{esc}=1.0 and f_{esc}=0.1, respectively. The galaxy contribution includes a redshift-dependent, *population-averaged* $f_{esc}^\star(z) \sim 25\% \rightarrow 5\%$ for star-forming galaxies from $z \sim 15 \rightarrow 4$, which was found to reproduce a set of IGM and CMB observational constraints on the reionization history (see section 3.2 and fig. 8 in Paper IV for a full discussion). These results are compared to observational constraints from Becker & Bolton (2013).

relations to the underlying physics. This approach also minimizes tracking of spatial elements, such as velocity and position, making it very efficient at sampling objects across a wide mass range at a fraction of the computational cost. An advantage of this approach over purely empirical models is that it is based on self-consistent underlying co-evolution of black holes and their galactic hosts according to physical laws. In addition, the highly flexible, modular structure of our modelling pipeline enables free exploration of different model configurations, including the choice of physical parameters and the recipe or prescriptions for specific physical processes.

5.1 High-redshift AGN populations and their role in cosmic reionization

The nature of the sources of reionization is one of the fundamental questions in cosmology. The interplay between AGN, which have harder spectra but are more rare, and galaxies, which are less efficient at producing ionizing photons, but are more common, is one of the primary open questions. Direct observational constraints on the faint AGN population at high redshift are not currently available. Therefore, simultaneously considering the observational constraints on hydrogen and helium reionization provides a promising avenue to address this issue. In the past, this has usually been addressed using empirical descriptions of the AGN luminosity function (extrapolated below the observed population) combined with simple scaling relations or template SEDs. In this work, we present a significant step forward by using a physics-based model, which we confirm produces AGN LFs that are consistent with observations within the uncertainties. This model is coupled with the physical AGN

spectral synthesis model of KD18, which we showed predicts considerably more scatter in key quantities than is generally reflected in the simplified scaling relations. Moreover, as these empirical approaches are typically based on bright (observable) AGN, we showed that they may systematically overestimate the ionizing photon budget.

5.1.1 Helium reionization

According to our predictions, if the AGN escape fraction were 100% (as assumed in many previous studies), helium would be ionized as early as $z \sim 5$. As this is inconsistent with observations which suggest that helium reionization is still in progress at $z \sim 2.7$, we suggest that a more likely scenario, and one that is consistent with our state-of-the-art knowledge about the fraction of obscured AGN and multi-wavelength AGN luminosity functions, is that the effective AGN escape fraction for helium ionizing photons is closer to 10% on average. Of course, how the effective escape fraction varies from one object to another, or effectively changes for the overall population over cosmic time, is extremely uncertain. We consider a scenario in which the escape fraction is as high as 30% at $z \sim 7$ and declines to 10% at $z \sim 2$. This results in the completion of helium reionization by $z \sim 3.5$, a bit early compared to some observational constraints. Our models predict that the bulk of the helium ionizing photons at $z > 4$ are from AGN with bolometric luminosities less than $10^{12} L_\odot$, and are hosted by halos with mass less than $10^{11.5} M_\odot$.

5.1.2 Hydrogen reionization

The other side of the coin is whether these AGN play a significant role in hydrogen reionization. To address this question, we consider the galaxy populations that are self-consistently predicted by these same models, along with their production and emission of ionizing photons, which have also been studied in Paper IV. We find that during the epoch of hydrogen reionization ($z > 6$), in our models AGN produce 1% or less of the ionizing photon budget relative to that produced by galaxies. Therefore, we conclude that AGN are unlikely to have contributed significantly throughout most of hydrogen reionization. However, the ionizing photon budget contributed by AGN rises rapidly at $z < 6$, with AGN contributing 30% of the ionizing photon budget at $z = 4$ and 60% at $z = 2$ (not accounting for any escape fraction). Thus, it appears that while the ionizing emissivity contributed by galaxies may be starting to decline already by $z \sim 4$, the contribution from AGN is rising steeply over this same interval, perhaps helping to account for the relatively constant ionizing emissivity implied by observations.

5.2 Our results in the context of other model predictions

We compared our predicted AGN luminosity functions with predictions from a compilation of state-of-the-art hydrodynamic simulations from $z = 2$–4 (predictions at higher redshifts were not easily available). It is interesting that these simulations produce a very wide range of predicted AGN LFs over this redshift range, with different normalizations in some cases, and in some cases very different faint end

slopes. Our predictions are by no means an outlier in this compilation, and in fact are roughly close to the median between the predictions from different hydrodynamic simulations. This highlights how observed AGN LFs at high redshift can help to constrain the physics regulating black hole accretion in current simulations.

We also anchored and cross checked our model outputs against a variety of observational constraints. For instance, the SEDs for AGN accreting at near-Eddington rates reproduces the observed power-law, as well as the correlation between $M_{\rm BH}$ and ϵ_{912}. We incorporate a detailed estimate of nuclear obscuration using the radiation lifted torus model of Buchner & Bauer (2017) to forward model our predictions for the hard X-ray luminosity function, and show that they agree well with observational constraints. We also compare our predictions with the estimated observed bolometric luminosity function reported by Shen et al. (2020), which they obtained by correcting for obscuration and attenuation and performing bolometric corrections on a comprehensive set of multi-wavelength AGN LFs, and find good agreement within the considerable uncertainties.

Although our predicted bolometric LFs are in good agreement with the observational constraints presented by S20, we predict a cosmic emissivity ϵ_{912} that is a factor of 10 higher than the one they quote over all redshifts (see Fig. 10). We have carried out a thorough comparison between our work and that of S20, and first ruled out the assumed AGN spectra as the main source of difference (see Fig. 3). Instead, the reason is that S20 calculated the emissivity at 1450Å (ϵ_{1450}) by integrating the observed UVLF and then assumed a power-law spectra to convert that into ϵ_{912}. Effectively, this assumes that obscured AGN, which are not present in UV-selected samples, do not contribute to the ionizing photon budget. We note that the other past studies compared in Fig. 10 also adopt a similar approach to track the ionizing emissivity and are rooted in similar UV1450 observations and empirical relations. However, we have pointed out that there is strong observational evidence that AGN obscuration and the escaping ionizing emission is extremely anisotropic. Therefore, although an AGN might drop out of the sample when observed from a particular line of sight, we would expect ionizing photons to potentially be able to escape over some solid angle in other directions. Therefore, this effective escape fraction (which likely depends on how AGN feedback has cleared gas and dust, producing channels where ionizing radiation can escape) is one of the key parameters in understanding the role of AGN in reionization. For this reason, we advocate for a cleaner separation between ionizing photon production rate and escape fraction. We suggest that the production of ionizing photon should first be estimated based on the unobscured, unattenuated luminosity function. This will have significantly different implications for physical quantities such as the growth rate and mass functions of SMBH at high redshift, which we may eventually be able to test through gravitational wave experiments such as LISA.

5.3   Major caveats and limitations of current models

In this subsection we discuss the most significant outstanding observational and theoretical uncertainties, including some caveats about our current models.

Similar to that of galaxies, the ionizing emissivity of AGN can be broken down into three moving parts, the abundance of AGN across cosmic time, their intrinsic production efficiency of ionizing photons, and the fraction of ionizing photons that escape to the IGM. The true abundance of AGN, especially the faint AGN that our models predict are the dominant source of ionizing photons in the early universe, are difficult or impossible to observe with current facilities. The deepest and most complete surveys are in the rest-UV, which may miss many objects either because they become too red to be included in colour-selected QSO surveys or because they become too faint to be detected at all. Enormous progress has been made in recent years in building up the X-ray detected AGN population at high redshift, but for the moment only the brightest objects can be detected at $z > 4$.

For the same reasons, the spectral characteristics of faint, high redshift AGN are also poorly constrained. Many studies adopt scaling relations between AGN bolometric luminosity and the luminosity in a specific band (bolometric corrections), which are obtained from nearby, luminous AGN. As we have pointed out, there can be scatter in these relations, which may correlate with the BH physical properties such as mass and accretion rate, and there are currently few observational constraints on the relevant populations. Even physics-based spectral models such as the one from KD18, used here, still contain some free parameters, which are currently calibrated to reproduce a handful of well-measured AGN spectra in the nearby universe. These parameters characterize the physical characteristics, such as electron temperature and spectral index for each of the radiating regions, as well as the radius and scale height of the disc region. Although these parameters are not anticipated to evolve as a function of redshift, some of these characteristics may have implicit dependences on the mass and accretion rate of the accreting black hole, and could vary to some degree for AGN of masses that have not been explicitly tested.

The ionizing photon escape fraction is the least constrained among the three moving parts. We are not aware of any direct observational constraints on the escape fraction of helium or hydrogen ionizing photons from AGN. Moreover, it is extremely difficult to obtain accurate predictions from simulations, since this quantity is certainly highly sensitive to the detailed structure of gas and dust over a vast range of scales, from the scales of the accretion disk and torus around the BH, to the galactic-scale outflows that may clear channels through which photons can escape (Benson et al. 2013; Seiler et al. 2018). This is a critical but challenging area for future work.

Turning to the theoretical models used in this work, we describe the main caveats and uncertainties.

We have used an Extended Press-Schechter based algorithm to generate merger trees, as this enabled us to probe a much larger dynamic range than would have been possible with N-body simulations. However, the EPS formalism is known to be inaccurate at the factor of 2 level, and is not well tested at extreme redshifts. We have previously compared our predictions for galaxy properties with predictions from other numerical models and simulations (Yung et al. 2019b; Behroozi et al. 2020), finding order unity agreement, providing confidence that this is not a significant source of error. EPS merger trees also do not correctly reflect the

known correlations, found in N-body simulations, between halo properties such as concentration, and mass accretion history, and do not enable predictions of spatial clustering of halos or their progenitors.

One of the major uncertainties in theoretical models of black hole formation is the mechanism or mechanisms by which seed black holes are formed in the early Universe, which has implications both for the masses of halos in which these objects form and their mass function. Scenarios include seeds from remnants of Pop III stars, which would populate nearly every halo down to very low masses ($\sim 10^5 M_\odot$) with 'light' seeds ($\sim 100 M_\odot$), to direct collapse or gravitational runaway scenarios, which would populate only a fraction of more massive halos with more massive seeds ($\sim 10^4 - 10^5 M_\odot$; for reviews, see (Volonteri 2010; Greene et al. 2020)). Most current cosmological simulations simply populate halos above a critical mass with a seed BH with a fixed mass. Our algorithm is similar: we populate every 'top level' halo with a seed BH with mass $m_{\rm seed}$. A heavy seeding mechanism could help to explain the difficulty of growing very massive black holes in the very early ($z > 7$) Universe without invoking super-Eddington accretion, but current theoretical predictions suggest that it is difficult to produce these seeds at a high enough number density to explain these objects (Visbal et al. 2018, 2020). For recent exploration of the impact of different seeding mechanisms on AGN populations at high redshift, see Ricarte & Natarajan (2018a), Ricarte & Natarajan (2018b).

A second major uncertainty is the intertwined processes that govern black hole accretion and feedback (which is thought to regulate the black hole accretion as well as star formation). Cosmological simulations typically assume a version of the Bondi-Hoyle model for BH accretion, but recent hyper-refinement simulations with direct capture to measure BH accretion have shown that this approximation can be off by many orders of magnitude, and may not even predict the correct dependence on BH mass (Anglés-Alcázar et al. 2020). Cosmological simulations also adopt various different phenomenological approximations to represent AGN feedback, often including multiple feedback 'modes' associated with rapid (close to Eddington) or slow (very sub-Eddington) accretion onto the BH (see Somerville & Davé 2015, and references therein for reviews). The large dispersion in the predicted AGN LFs from different simulations shown here presumably largely reflects the major uncertainties in modeling these processes. In this work, we adopt the model for black hole growth and feedback first presented in S08, and extensively tested in the context of predictions for AGN populations in Hirschmann et al. (2012). Although this model is based on a generation of hydrodynamic simulations that is no longer state-of-the-art, we argue that our predictions are still robust for this study, both because we have demonstrated good agreement with observed AGN LFs and because our predictions are consistent with those of recent cosmological hydrodynamic simulations. However, more physical modeling of these processes is clearly a high priority for the next generation of models.

Finally, we adopted a simplified analytic model to estimate the volume filling fraction of ionized hydrogen and helium. Similar to our discussion on the reionization model in Paper IV, this analytic approach does not capture the potential 'patchiness' of the progression in the evolution of ion-

ized volume due to the density fluctuations and clustering of sources, as shown by numerous numerical simulations (Trac & Cen 2007; Hassan et al. 2016, 2017; Mutch et al. 2016; Hutter et al. 2020). We add that while the epochs of hydrogen and helium reionization are treated as separate events in this work, which is a practice that has been justified and adopted in many similar studies, in reality, the presence of early X-ray sources may heat up both both hydrogen and helium in the IGM, giving rise to an additional channel where intergalactic hydrogen can be collisionally ionized by free electrons arising from photoionized He I (Venkatesan et al. 2001). This effect will be further explored in an upcoming work where we fully couple the AGN–helium reionization framework presented in this work and the galaxy–hydrogen reionization framework presented in Paper IV (Yung et al., in preparation).

In addition, it is evident that AGN emission is highly stochastic and anisotropic (Lau et al. 2017), which causes additional complications for detailed radiation transport modeling. As the IGM gas density in highly clustered regions is expected to be higher than the universal average, the shorter recombination timescale will cause a slow down as the ionization front progresses through these regions. As it is known that luminous quasars tend to form in more highly clustered regions than low-mass galaxies, this could have important implications for the relative topology of hydrogen vs. helium, which however our current models cannot capture.

5.4 Outlook for probing faint high-z AGN with *JWST* and beyond

One of the main science goals of the highly-anticipated *JWST* is to survey the extremely distant universe, and it is expected that *JWST* will detect objects several orders of magnitude below the current limit of *HST*. Photometric and grism surveys with the on-board Near-Infrared Camera (NIRCam) and Near-Infrared Imager and Slitless Spectrograph (NIRISS) will detect large numbers of objects at high redshift. While these upcoming surveys are predominantly searching for star-forming galaxies, they will also presumably detect some AGN. Follow-up spectroscopy with the Near-Infrared Spectrograph (NIRSpec) will enable identification of AGN candidates and will provide insights about their spectral characteristics. Particularly promising is the use of nebular emission line ratios in the rest-UV, which have been shown to provide a more robust discriminator between AGN and SF/composite galaxies at high redshift than traditional BPT diagnostics (Feltre et al. 2016; Hirschmann et al. 2017, 2019). We intend to extend our models in the near future to enable predictions of such emission lines and explore the potential of *JWST* to identify faint, high-redshift AGN in this manner. Moreover, we expect the extended line emission predictions will deliver additional observable signatures that enable direct comparison with low-redshift analogs, such as Green Pea galaxies (e.g. Jaskot et al. 2019).

Planned wide-field survey instruments, such as the *Nancy Grace Roman Space Telescope* (Spergel et al. 2015, formerly known as the *Wide-Field Infrared Survey Telescope* or *WFIRST*) and the *Vera C. Rubin Observatory* (LSST Science Collaboration 2017, formerly the *Large Synoptic Survey Telescope* or *LSST*) will increase survey coverage by a factor of hundreds compared to current generation instru-

ments, which will dramatically increase the ability to find these rare objects. In the near future, we plan to deliver forecasts in mock lightcone format similar to the ones presented by Somerville et al. (2021) but of multiple square degrees (Yung et al., in preparation).

Looking further into the future, upcoming *approved* facilities, such as the X-ray *Advanced Telescope for High ENergy Astrophysics* (*ATHENA*) mission and the gravitational wave detector *Laser Interferometer Space Antenna* (*LISA*), as well as the proposed NASA X-ray mission concepts Advanced X-ray Imaging Satellite (AXIS) and Lynx (Weisskopf et al. 2015), will enable multi-wavelength and multi-messenger surveys that will provide unprecedented constraints on the high redshift population of supermassive black holes and their accretion rates.

6   SUMMARY AND CONCLUSIONS

In this work, we constructed a computationally efficient modelling pipeline that is based on the versatile Santa Cruz SAM. Benefiting from the wide dynamic range achievable by the SAM, we are able to self-consistently model a wide range of objects, including the ones that are too rare to be captured in cosmological hydrodynamic simulations and the ones that are too faint to be seen in current observations. This is achieved by coupling the SAM with the KD18 physical AGN spectral model, which outputs physically-motivated SEDs from a model with three distinct radiating zones near SMBHs, dependent on black hole mass and accretion rate, with a number of assumptions regarding the geometry and conditions within these zones. By combining the SAM prediction for the distributions of BH masses and AGN accretion rates at different cosmic times with these SED models, we compute the cosmic budget for both hydrogen and helium ionizing photons. This is then combined with an analytic reionization model to forward model the cosmic reionization history.

By making use of three physically motivated models that operate at vastly different physical scales, we connect the 'ground-level', small-scale black hole physics to the 'top-level' cosmological observables, such as AGN luminosity functions and the IGM reionization history across cosmic time. This gives a comprehensive *overview* of the set of physics derived from observing the nearby Universe and simulations, and provides an *outlook* on the objects forming in the high-redshift Universe to be detected with the next generation of instruments.

We presented predictions for AGN populations at $z = 2$ to 7 with the black hole growth feedback models embedded in the well-established Santa Cruz SAM. We found that the recently published version of these models from Somerville et al. (2021) and the Yung et al. 'semi-analytic forecasts for JWST' series produce very good agreement with the hard X-ray luminosity function.

We conducted a controlled experiment and found that increasing the scatter in the $m_{\rm BH}$–$m_{\rm bulge}$ relation from the value calibrated at $z \sim 0$ yields more luminous AGN at $z \gtrsim 2$ and improves the agreement with the bright end of the observed AGN luminosity functions between $z \sim 2$ to 4.

We constructed an extended modelling pipeline to link the predicted AGN populations to the subsequent reioniza-

tion history of intergalactic hydrogen and helium. This is done by forward modelling the AGN SEDs based on the predicted black hole mass and accretion rate in halos across a wide mass range. We provided predictions for the helium ionizing emissivity by summing over the contribution from all predicted AGN. We then tracked the progression of the ionized volume fraction for intergalactic helium utilizing an analytic reionization model. Our models enable self-consistent investigation of the evolution in the AGN populations across cosmic time and their impacts on the cosmic environment.

This work adds new capabilities of predicting spectra and photometry for AGN, as well as their underlying physical properties to a larger *forecasting* framework, which provides tailored predictions for high-redshift observations with *JWST* and other instruments.

We summarize our main conclusions below:

(i) Using an AGN model that is empirically calibrated and well-tested at low redshifts in conjunction with a physical AGN spectral model, we provide predictions for AGN LFs, helium ionizing photon production rate, and ionized volume fraction between $z = 2$ to 7.

(ii) We show that empirical relations that are commonly used to link different observable quantities neglect significant scatter that arises from variation in the physical properties of AGN.

(iii) Our physical model predicts a factor of ten higher helium ionizing emissivity compared to many past studies. We show that this is due to the assumption in previous studies that AGN that are missing from observed UV luminosity functions due to attenuation or obscuration do not contribute to the ionizing photon budget.

(iv) Our self-consistent calculations demonstrate that hard ionizing photons produced by accreting SMBHs between $z = 2$ to 7 are sufficient to fully reionize the intergalactic helium by $z \sim 2$, with a moderate assumed escape fraction of $\sim 10\%$. This roughly agrees with the fraction of Compton-thin, UV-unattenuated AGN estimated from AGN surveys.

(v) According to our predictions, AGN contribute less than ten percent of the hydrogen ionizing photons at $z > 6$.

(vi) We provide scaling relations for $\dot{N}_{\rm ion, He\,II}$ as a function of $M_{\rm h}$ throughout the epoch of helium reionization, which may be useful for cosmological-scale reionization simulations where baryonic physics is not resolved.

(vii) We project that AGN of $\log L_{\rm bol}/{\rm L}_\odot = 10$ to 11 are responsible for producing most of the ionizing photons at $z \sim 3$ to 4, during the epoch where intergalactic helium is getting reionized most rapidly.

(viii) The contribution of ionizing photons from AGN increases rapidly at $z < 6$, until they are contributing an equal or slightly larger amount than star forming galaxies by $z \sim 2$. This may help to sustain hydrogen ionization at late times, in a scenario where galaxy escape fractions turn out to be extremely low (less than a few percent).

DATA AVAILABILITY

The data underlying this paper are available in the Data Product Portal hosted by the Flatiron Institute at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/.

REFERENCES

Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93
Aird J., Coil A. L., Georgakakis A., Nandra K., Barro G., Pérez-González P. G., 2015a, MNRAS, 451, 1892
Aird J., et al., 2015b, ApJ, 815, 66
Akiyama M., et al., 2018, Publ. Astron. Soc. Japan, 70, 1
Amarantidis S., et al., 2019, MNRAS, 485, 2694
Anglés-Alcázar D., Faucher-Giguère C. A., Quataert E., Hopkins P. F., Feldmann R., Torrey P., Wetzel A., Kereš D., 2017, MNRAS, 472, L109
Anglés-Alcázar D., et al., 2020, arXiv:2008.12303
Arnaud K. A., 1996, in Jacoby G. H., Barnes J., eds, Vol. 101, Astron. Data Anal. Softw. Syst. V, A.S.P. Conf. Ser.. p. 17
Assef R. J., et al., 2011, ApJ, 728
Becker G. D., Bolton J. S., 2013, MNRAS, 436, 1023
Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096
Behroozi P., et al., 2020, MNRAS, 499, 5702
Benson A., Venkatesan A., Shull J. M., 2013, ApJ, 770, 76
Bohlin R. C., Savage B. D., Drake J. F., 1978, ApJ, 224, 132
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bondi H., 1952, MNRAS, 112, 195
Bongiorno A., et al., 2007, A&A, 472, 443
Bovy J., et al., 2011, ApJ, 729
Brennan R., et al., 2015, MNRAS, 451, 2933
Buchner J., Bauer F. E., 2017, MNRAS, 465, 4348
Buchner J., et al., 2015, ApJ, 802, 89
Choi E., Naab T., Ostriker J. P., Johansson P. H., Moster B. P., 2014, MNRAS, 442, 440
Choi E., Ostriker J. P., Naab T., Somerville R. S., Hirschmann M., Núñez A., Hu C.-Y., Oser L., 2017, ApJ, 844, 31
Choudhury T. R., Ferrara A., 2005, MNRAS, 361, 577
Civano F., et al., 2016, ApJ, 819, 62
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, MNRAS, 486, 2827
Dawson K. S., et al., 2013, AJ, 145
Dayal P., et al., 2020, MNRAS, 495, 3065

Dijkstra M., Gronke M., Venkatesan A., 2016, ApJ, 828, 71
Done C., Gierliński M., Kubota A., 2007, Astron. Astrophys. Rev., 15, 1
Done C., Davis S. W., Jin C., Blaes O., Ward M., 2012, MNRAS, 420, 1848
Draine B. T., 2003, ARA&A, 41, 241
Dubois Y., et al., 2014, MNRAS, 444, 1453
Dubois Y., Peirani S., Pichon C., Devriendt J., Gavazzi R., Welker C., Volonteri M., 2016, MNRAS, 463, 3948
Ebrero J., et al., 2009, A&A, 493, 55
Efstathiou G., Lake G., Negroponte J., 1982, MNRAS, 199, 1069
Fan X., Carilli C., Keating B., 2006a, ARA&A, 44, 415
Fan X., et al., 2006b, AJ, 132, 117
Fanidakis N., et al., 2013, MNRAS, 435, 679
Faucher-Giguère C.-A. A., 2021, MNRAS, 493, 1614
Feltre A., Charlot S., Gutkin J., 2016, MNRAS, 456, 3354
Finkelstein S. L., 2016, Publ. Astron. Soc. Aust., 33, e037
Finkelstein S. L., et al., 2019, ApJ, 879, 36
Fletcher T. J., Tang M., Robertson B. E., Nakajima K., Ellis R. S., Stark D. P., Inoue A., 2019, ApJ, 878, 87
Fontanot F., et al., 2020, MNRAS, 496, 3943
Gardner J. P., et al., 2006, Space Sci. Rev., 123, 485
Gaskell C. M., Benker A. J., 2007, arXiv:0711.1013
Genel S., et al., 2014, MNRAS, 445, 175
Giallongo E., et al., 2015, A&A, 578, A83
Glikman E., Djorgovski S. G., Stern D., Dey A., Jannuzi B. T., Lee K. S., 2011, ApJL, 728, 1
Greene J. E., Strader J., Ho L. C., 2020, ARA&A, 58, 257
Grogin N. A., et al., 2011, ApJS, 197, 35
Guaita L., et al., 2016, A&A, 587, A133
Haardt F., Madau P., 2012, ApJ, 746, 125
Habouzit M., et al., 2021, MNRAS, 503, 1940
Hasinger G., Miyaji T., Schmidt M., 2005, A&A, 441, 417
Hassan S., Davé R., Finlator K., Santos M. G., 2016, MNRAS, 457, 1550
Hassan S., Davé R., Finlator K., Santos M. G., 2017, MNRAS, 468, 122
Hassan S., Davé R., Mitra S., Finlator K., Ciardi B., Santos M. G., 2018, MNRAS, 473, 227
Heap S. R., Williger G. M., Smette A., Hubeny I., Sahu M. S., Jenkins E. B., Tripp T. M., Winkler J. N., 2000, ApJ, 534, 69
Hirschmann M., Somerville R. S., Naab T., Burkert A., 2012, MNRAS, 426, 237
Hirschmann M., Dolag K., Saro A., Bachmann L., Borgani S., Burkert A., 2014, MNRAS, 442, 2304
Hirschmann M., Charlot S., Feltre A., Naab T., Choi E., Ostriker J. P., Somerville R. S., 2017, MNRAS, 472, 2468
Hirschmann M., Charlot S., Feltre A., Naab T., Somerville R. S., Choi E., 2019, MNRAS, 487, 333
Hiss H., Walther M., Hennawi J. F., Oñorbe J., O'Meara J. M., Rorai A., Lukić Z., 2018, ApJ, 865, 42
Hopkins P. F., Quataert E., 2010, MNRAS, 407, 1529
Hopkins P. F., Richards G. T., Hernquist L., 2007a, ApJ, 654, 731
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007b, ApJ, 669, 45
Hopkins P. F., Hernquist L., Cox T. J., Robertson B., Krause E., 2007c, ApJ, 669, 67
Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27
Hutter A., Dayal P., Yepes G., Gottlöber S., Legrand L., Ucci G., 2020, arXiv:2004.08401
Jakobsen P., Boksenberg A., Deharveng J., Greenfield P., Jedrzejewski R., Paresce F., 1994, Nature, 370, 35
Jansen F., et al., 2001, A&A, 365, L1
Jaskot A. E., Dowd T., Oey M. S., Scarlata C., McKinney J., 2019, ApJ, 885, 96
Jiang F., van den Bosch F. C., 2014, MNRAS, 440, 193
Jiang L., et al., 2008, AJ, 135, 1057

Jin C., Ward M., Done C., Gelbord J., 2012a, MNRAS, 420, 1825

Jin C., Ward M., Done C., 2012b, MNRAS, 422, 3268

Kaviraj S., et al., 2017, MNRAS, 467, stx126

Khorunzhev G. A., Sazonov S. Y., Burenin R. A., 2018, AL, 44, 500

Kimm T., Cen R., 2014, ApJ, 788, 121

Kimm T., Katz H., Haehnelt M., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 466, 4826

Kimm T., Blaizot J., Garel T., Michel-Dansac L., Katz H., Rosdahl J., Verhamme A., Haehnelt M., 2019, MNRAS, 486, 2215

Klypin A., Yepes G., Gottlöber S., Prada F., Heß S., 2016, MNRAS, 457, 4340

Kocevski D. D., et al., 2017, ApJ, 846, 112

Koekemoer A. M., et al., 2011, ApJS, 197, 36

Krawczyk C. M., Richards G. T., Mehta S. S., Vogeley M. S., Gallagher S. C., Leighly K. M., Ross N. P., Schneider D. P., 2013, ApJS, 206, 4

Kubota A., Done C., 2018, MNRAS, 480, 1247

Kulkarni G., Worseck G., Hennawi J. F., 2019, MNRAS, 488, 1035

LSST Science Collaboration 2017, arXiv:1708.04058

Lacey C., Cole S., 1993, MNRAS, 262, 627

Lau M. W., Xavier Prochaska J., Hennawi J. F., 2017, arXiv, 857, 126

Luo B., et al., 2016, ApJS, 228, 2

Lusso E., Worseck G., Hennawi J. F., Prochaska J. X., Vignali C., Stern J., O'Meara J. M., 2015, MNRAS, 449, 4204

Ma X., et al., 2015, MNRAS, 453, 960

Ma X., Quataert E., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2020, MNRAS, 498, 2001

Madau P., Dickinson M., 2014, ARA&A, 52, 415

Madau P., Haardt F., 2015, ApJ, 813, L8

Madau P., Haardt F., Rees M. J., 1999, ApJ, 514, 648

Manti S., Gallerani S., Ferrara A., Greig B., Feruglio C., 2017, MNRAS, 466, 1160

Marconi A., Risaliti G., Gilli R., Hunt L. K., Maiolino R., Salvati M., 2004, MNRAS, 351, 169

Masters D., et al., 2012, ApJ, 755, 169

Matsuoka Y., et al., 2018, ApJ, 869, 150

Matsuoka Y., et al., 2019, ApJ, 883, 183

McConnell N. J., Ma C.-P., 2013, ApJ, 764, 184

McGreer I. D., et al., 2013, ApJ, 768, 105

McGreer I. D., Fan X., Jiang L., Cai Z., 2018, AJ, 155, 131

McQuinn M., Lidz A., Zaldarriaga M., Hernquist L., Hopkins P. F., Dutta S., Faucher-Giguère C. A., 2009, ApJ, 694, 842

Mehdipour M., et al., 2011, A&A, 534, 1

Mehdipour M., et al., 2015, A&A, 575, 1

Menci N., Gatti M., Fiore F., Lamastra A., 2014, A&A, 569

Miyaji T., et al., 2015, ApJ, 804, 104

Morrison R., McCammon D., 1983, ApJ, 270, 119

Mutch S. J., Geil P. M., Poole G. B., Angel P. W., Duffy A. R., Mesinger A., Wyithe J. S. B., 2016, MNRAS, 462, 250

Nakajima K., Ellis R. S., Robertson B. E., Tang M., Stark D. P., 2020, ApJ, 889, 161

Nandra K., et al., 2015, Astrophys. Journal, Suppl. Ser., 220, 10

Natarajan P., Pacucci F., Ferrara A., Agarwal B., Ricarte A., Zackrisson E., Cappelluti N., 2017, ApJ, 838, 117

Nelson D., et al., 2019, Comput. Astrophys. Cosmol., 6, 2

Novikov I. D., Thorne K. S., 1973, in DeWitt C., DeWitt B., eds, , Black holes. Gordon and Breach, New York, p. 343

Nulsen P. E., Fabian A. C., 2000, MNRAS, 311, 346

Orofino M. C., Ferrara A., Gallerani S., 2021, arXiv:2101.05292

Paardekooper J.-P., Pelupessy F. I., Altay G., Kruip C., 2011, A&A, 530, A87

Paardekooper J.-P., Khochfar S., Dalla Vecchia C., 2015, MNRAS, 451, 2544

Palanque-Delabrouille N., et al., 2013, A&A, 551, 1

Palanque-Delabrouille N., et al., 2016, A&A, 587, A41

Pawlik A. H., Schaye J., Vecchia C. D., 2015, MNRAS, 451, 1586

Pei Y. C., 1992, ApJ, 395, 130

Peterson P., 2009, Int. J. Comput. Sci. Eng., 4, 296

Planck Collaboration 2016a, A&A, 594, A13

Planck Collaboration 2016b, A&A, 596, A108

Popping G., Somerville R. S., Trager S. C., 2014, MNRAS, 442, 2398

Popping G., Puglisi A., Norman C. A., 2017, MNRAS, 472, 2315

Porter L. A., Somerville R. S., Primack J. R., Johansson P. H., 2014, MNRAS, 444, 942

Press W. H., Schechter P., 1974, ApJ, 187, 425

Price-Whelan A. M., et al., 2018, AJ, 156, 123

Puchwein E., Bolton J. S., Haehnelt M. G., Madau P., Becker G. D., Haardt F., 2015, MNRAS, 450, 4081

Ricarte A., Natarajan P., 2018a, MNRAS, 474, 1995

Ricarte A., Natarajan P., 2018b, MNRAS, 481, 3278

Richards G. T., et al., 2002, AJ, 123, 2945

Robitaille T. P., et al., 2013, A&A, 558, A33

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Ross N. P., et al., 2013, ApJ, 773, 14

Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L., 2000, MNRAS, 318, 817

Schaye J., et al., 2015, MNRAS, 446, 521

Schirber M., Bullock J. S., 2003, ApJ, 584, 110

Seiler J., Hutter A., Sinha M., Croton D., 2018, MNRAS, 480, L33

Shapley A. E., Steidel C. C., Strom A. L., Bogosavljević M., Reddy N. A., Siana B., Mostardi R. E., Rudie G. C., 2016, ApJ, 826, L24

Shen Y., Kelly B. C., 2012, ApJ, 746

Shen X., Hopkins P. F., Faucher-Giguère C.-A., Alexander D. M., Richards G. T., Ross N. P., Hickox R. C., 2020, MNRAS, 495, 3252

Sheth R. K., Tormen G., 2002, MNRAS, 329, 61

Shull J. M., France K., Danforth C. W., Smith B., Tumlinson J., 2010, ApJ, 722, 1312

Singal J., George J., Gerber A., 2016, ApJ, 831, 60

Somerville R. S., 2009, MNRAS, 399, 1988

Somerville R. S., Davé R., 2015, ARA&A, 53, 31

Somerville R. S., Kolatt T. S., 1999, MNRAS, 305, 1

Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087

Somerville R. S., Primack J. R., Faber S. M., 2001, MNRAS, 320, 504

Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481

Somerville R. S., Gilmore R. C., Primack J. R., Domínguez A., 2012, MNRAS, 423, 1992

Somerville R. S., Popping G., Trager S. C., 2015, MNRAS, 453, 4338

Somerville R. S., et al., 2021, MNRAS, 502, 4858

Spergel D., et al., 2015, arXiv:1503.03757

Su K.-Y., et al., 2021, arXiv:2102.02206

Syphers D., Shull J. M., 2013, ApJ, 765, 119

Syphers D., Shull J. M., 2014, ApJ, 784

Telfer R. C., Zheng W., Kriss G. A., Davidsen A. F., 2002, ApJ, 565, 773

Thomas N., Davé R., Jarvis M. J., Anglés-Alcázar D., 2021, MNRAS, 503, 3492

Toomre A., 1964, ApJ, 139, 1217

Trac H., Cen R., 2007, ApJ, 671, 1

Trebitsch M., Blaizot J., Rosdahl J., Devriendt J., Slyz A., 2017, MNRAS, 470, 224

Trump J. R., et al., 2011, ApJ, 733, 60

Ueda Y., Akiyama M., Ohta K., Miyaji T., 2003, ApJ, 598, 886

Ueda Y., Akiyama M., Hasinger G., Miyaji T., Watson M. G., 2014, ApJ, 786

Upton Sanderbeck P. R., D'Aloisio A., McQuinn M. J., 2016, MNRAS, 460, 1885

Venkatesan A., Giroux M. L., Shull J. M., 2001, ApJ, 563, 1

Virtanen P., et al., 2020, Nat. Methods, 17, 261

Visbal E., Haiman Z., Bryan G. L., 2018, MNRAS, 475, 5246

Visbal E., Bryan G. L., Haiman Z., 2020, ApJ, 897, 95

Vogelsberger M., et al., 2014, MNRAS, 444, 1518

Volonteri M., 2010, Astron. Astrophys. Rev., 18, 279

Volonteri M., Dubois Y., Pichon C., Devriendt J., 2016, MNRAS, 460, 2979

Volonteri M., Reines A. E., Atek H., Stark D. P., Trebitsch M., 2017, ApJ, 849, 155

Volonteri M., Pfister H., Beckmann R. S., Dubois Y., 2020, arXiv:2005.04902

Wang F., et al., 2019, ApJ, 884, 30

Weinberger R., et al., 2017, MNRAS, 465, 3291

Weisskopf M. C., Tananbaum H. D., Van Speybroeck L. P., O'Dell S. L., 2000, X-Ray Opt. Instruments, Mission. III, 4012, 2

Weisskopf M. C., Gaskin J., Tananbaum H., Vikhlinin A., 2015, EUV and X-ray Optics: Synergy between Laboratory and Space IV, 9510, 951002

Wyithe J. S. B., Loeb A., 2003, ApJ, 586, 693

Xu H., Wise J. H., Norman M. L., Ahn K., O'Shea B. W., 2016, ApJ, 833, 84

Xue Y. Q., Luo B., Brandt W. N., Alexander D. M., Bauer F. E., Lehmer B. D., Yang G., 2016, ApJS, 224, 15

Yang J., et al., 2018, AJ, 155, 110

Yang G., et al., 2020, arXiv:2011.08192

York D. G., et al., 2000, AJ, 120, 1579

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., 2019a, MNRAS, 483, 2983

Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., Ferguson H. C., Davé R., 2019b, MNRAS, 490, 2855

Yung L. Y. A., Somerville R. S., Popping G., Finkelstein S. L., 2020a, MNRAS, 494, 1002

Yung L. Y. A., Somerville R. S., Finkelstein S. L., Popping G., Davé R., Venkatesan A., Behroozi P., Ferguson H. C., 2020b, MNRAS, 496, 4574

Yung L. Y. A., et al., 2021, in preperation

Zdziarski A. A., Johnson W. N., Magdziarz P., 1996, MNRAS, 283, 193

Zhang J., Fakhouri O., Ma C. P., 2008, MNRAS, 389, 1521

Zycki P. T., Done C., Smith D. A., 1999, MNRAS, 309, 561

van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22

APPENDIX A: AGN UV ATTENUATION AND OBSCURATION

In this Appendix, we provide an additional diagnostic diagram, Fig. A1, for UV attenuation applied to our AGN based on the prescriptions described in Section 3.1.2. This relation is not expected to evolve with time as none of the model components involved in this calculation explicitly depends on redshift. We also note that the upper limit of this relation is inherited from the fixed ratio that converts neutral hydrogen to colour excess. The data points in the scatter plot are colour-coded for the X-ray luminosity L_x of predicted AGN, which is one of the variables that influences the prediction of N_H in the radiation-lifted torus model proposed by Buchner & Bauer (2017). This figure is comparable to a similar figures for galaxies presented in fig. 4 in Paper I.
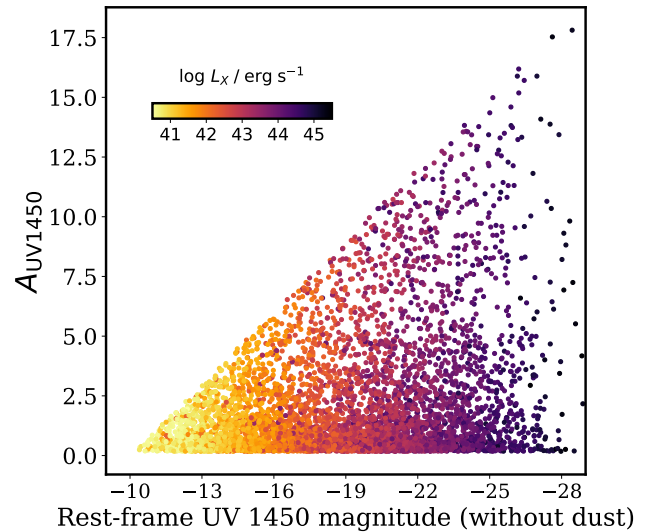
Figure A1. Distribution of intrinsic rest-frame UV1450 magnitude versus UV-band attenuation for our simulated AGN population. Here we show all AGN predicted across $z = 2 - 7$ since none of the model components used in this calculation explicitly depend on redshift. The data points are colour-coded to indicate their intrinsic hard X-ray luminosity.

APPENDIX B: EXTENDED PREDICTIONS ON UV LUMINOSITY FUNCTIONS

Upon updating the black hole growth model to include the feeding of cold gas via disc instability, we are also predicting some stronger AGN feedback onto the galaxies forming in larger halos. As a result of that, fewer bright, massive galaxies have formed, and therefore the amount of attenuation by dust required to reproduce the observed UV LFs have gone down slightly. The re-calibrated redshift evolution of the redshift-dependent dust optical depth is $\tau_{\mathrm{dust},0} = -0.001543z + 0.01583$. See section 2.6 in Paper I for full description for our dust attenuation model. We compare this updated parameter with previous work in Fig. B1 (Somerville et al. 2012; Yung et al. 2019a).

We present rest-frame UV (1600Å) luminosity functions at $z = 4$ to 7 from the set of predictions shown in this work (red, forming in root halos spanning $V_{\mathrm{vir}} = 100 - 1400\,\mathrm{km\,s}^{-1}$), compared to predictions from previous work (blue, $V_{\mathrm{vir}} = 20 - 500\,\mathrm{km\,s}^{-1}$). These luminosities only account for radiation from stellar populations but not AGNs. Both sets of predictions share an identical 'fiducial' model configurations (see Paper I for detail). The finer grid of root halos behind these new predictions provides more reliable sampling for the population of bright galaxies forming in massive halos. Note that massive objects at this redshift range are extremely rare and they are unlikely to be picked up by the relatively small field of view of *JWST*. These results are shown in Fig. B2 and tabulated values in Table B1.
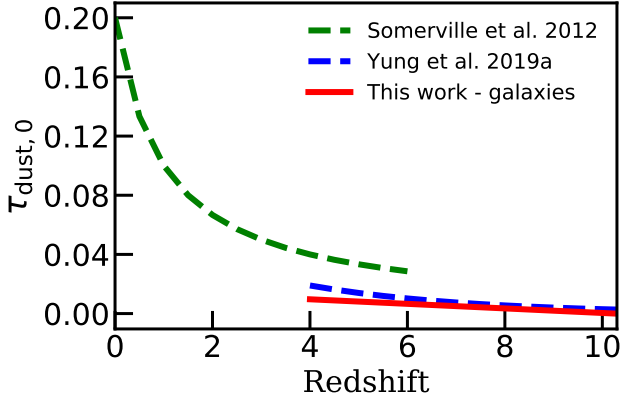
Figure B1. The $\tau_{\mathrm{dust},0}(z)$ values required to reproduce observations in the current model (red solid line), compared to the values adopted in Somerville et al. (2012, green dashed line) and Yung et al. (2019a, blue dashed line).

Table B1. Tabulated UV LFs at $z = 4$–7 predicted for new halo mass range as presented in Fig. B2.

M_{UV}	$z = 4$	$z = 5$	$z = 6$	$z = 7$
-25.5	-11.97	—	—	—
-25.0	-9.73	-12.78	—	—
-24.5	-9.22	-10.53	-11.03	-12.08
-24.0	-8.18	-8.86	-10.20	-10.44
-23.5	-6.87	-7.66	-8.33	-8.92
-23.0	-6.12	-6.78	-6.98	-7.25
-22.5	-5.12	-5.35	-5.75	-5.91
-22.0	-4.36	-4.46	-4.81	-5.06

APPENDIX C: COMPARISON TO EMPIRICAL CONVERSION

In Fig. C1, we compare the radiative efficiency $\epsilon_r \equiv L_{\mathrm{bol}}/(\dot{M}_{\mathrm{acc}}c^2)$ from our predictions with the typical values adopted in past studies (Hopkins et al. 2007a; Hirschmann et al. 2012). While some $z \sim 0$ observations indicate that an acceptable range would be $0.04 < \epsilon_r < 0.16$ (Marconi et al. 2004), this work demonstrates how this quantity varies as a function of BH mass and accretion rate.

APPENDIX D: SELECTED TABULATED DATA

We provide tabulated data for AGN LFs and ionizing emissivity as a function of halo mass in Table D1 and Table D2. The full range of tabulated data is available online in our data release portal.

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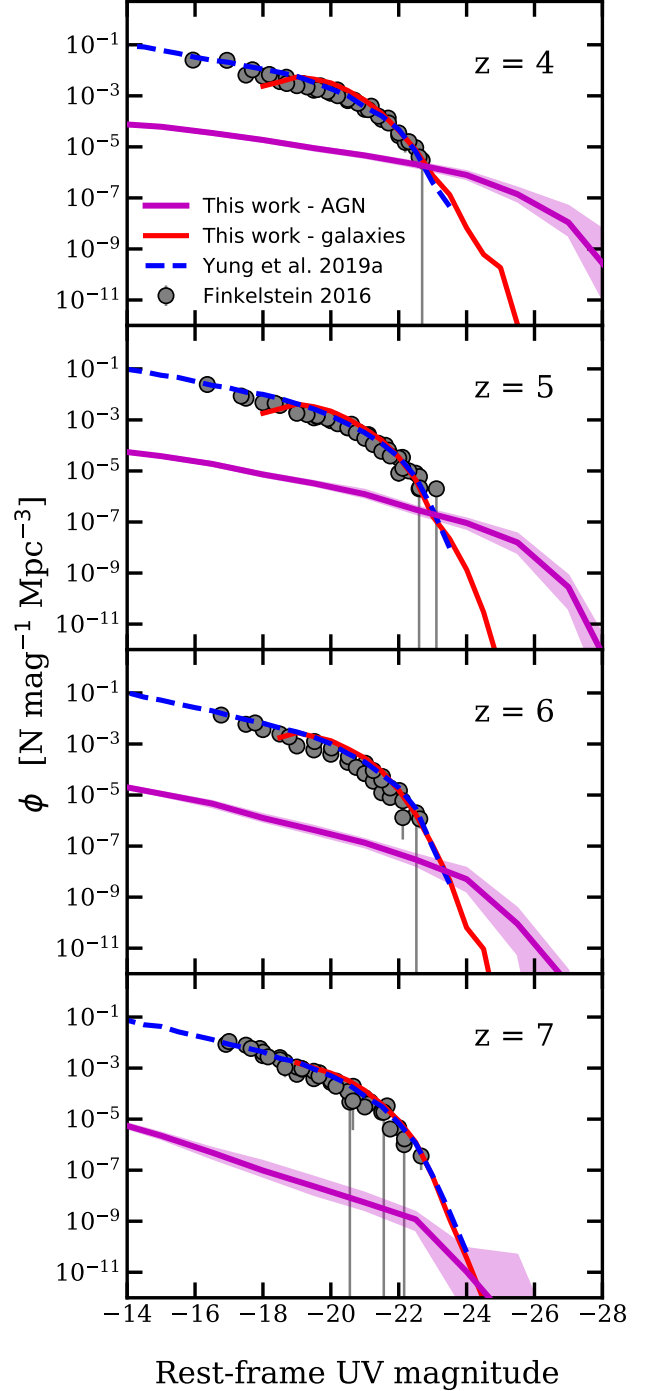

Figure B2. Rest-frame UV LFs from our fiducial model for galaxies predicted in this work (red, $V_{\mathrm{vir}} = 100 - 1400\,\mathrm{km\,s^{-1}}$) and from previous works in the paper series (blue, $V_{\mathrm{vir}} = 20 - 500\,\mathrm{km\,s^{-1}}$). We also include a compilation of observational constraints from Finkelstein (2016, circles) as shown in Paper I. We also include the AGN UVLFs predicted in this work for comparison (magenta line and shaded area, see Fig. 7 and associated text for detail).
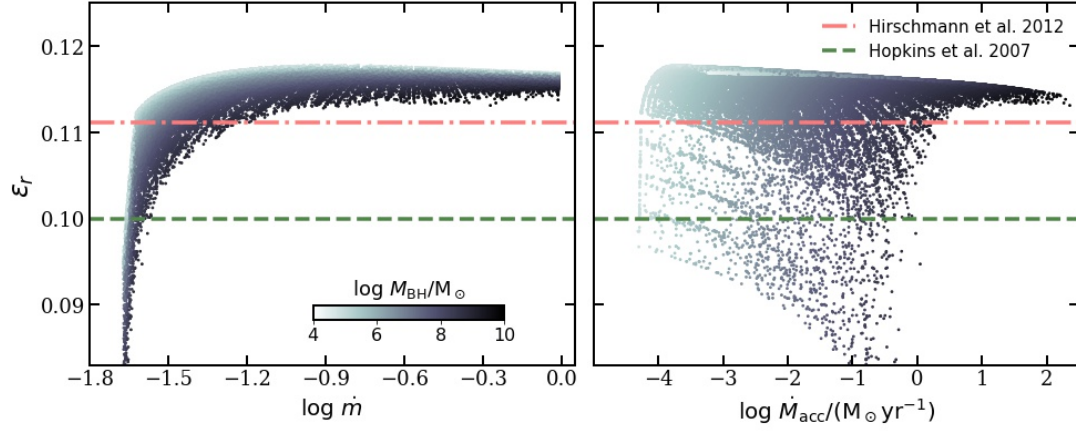
Figure C1. *Left panel:* AGN radiative efficiency ϵ_r versus Eddington-normalized accretion rate, $\dot{m}$, colour-coded for $M_{\rm BH}$, assuming isotropic emission (inclination $45°$). This distribution is compared to the constant values adopted in Hopkins et al. (2007a) and Hirschmann et al. (2012). *Right panel:* ϵ_r versus accretion rate in physical units, colour-coded for $M_{\rm BH}$. These figures illustrate the large scatter in radiative efficiency that is not captured by commonly used empirical conversions.

Table D1. Tabulated AGN LFs at $z = 2$–7 predicted for new halo mass range.

$\log(L_{\rm bol}/{\rm L_\odot})$	$z = 2$	$z = 3$	$z = 4$	$z = 5$	$z = 6$	$z = 7$
9.0	−3.61	−3.05	−2.81	−2.78	−3.03	−3.29
9.5	−3.32	−2.91	−2.78	−2.89	−3.21	−3.65
10.0	−3.25	−3.01	−2.97	−3.11	−3.54	−4.19
10.5	−3.45	−3.22	−3.23	−3.52	−4.05	−4.91
11.0	−3.65	−3.47	−3.55	−3.89	−4.58	−5.53
11.5	−3.85	−3.75	−3.87	−4.26	−5.30	−6.31
12.0	−4.18	−4.04	−4.23	−4.81	−5.65	−7.22
12.5	−4.47	−4.42	−4.59	−5.41	−6.35	−7.76
13.0	−4.74	−4.82	−5.03	−5.87	−7.61	−9.33
13.5	−5.21	−5.26	−5.80	−6.66	−8.61	—
14.0	−5.72	−6.36	−6.96	−8.38	—	—
14.5	−6.38	−7.58	−8.03	—	—	—
15.0	−8.89	−8.39	—	—	—	—

Table D2. Tabulated $\log(\dot{N}_{\rm ion,He}/M_{\rm h})$ at $z = 2$–7 predicted for new halo mass range.

$\log(M_{\rm h}/{\rm M}_\odot)$	$\log((\dot{N}_{\rm ion,He}/M_{\rm h})/({\rm s}\ {\rm M}_\odot)^{-1})$					
	$z = 2$	$z = 3$	$z = 4$	$z = 5$	$z = 6$	$z = 7$
11.00	$52.47^{+0.75}_{-0.80}$	$52.20^{+0.85}_{-0.86}$	$52.01^{+0.83}_{-0.81}$	$52.06^{+0.76}_{-0.85}$	$52.11^{+0.70}_{-0.78}$	$52.08^{+0.60}_{-0.58}$
11.50	$52.82^{+0.99}_{-0.95}$	$52.78^{+1.07}_{-1.04}$	$52.61^{+1.08}_{-1.09}$	$52.57^{+0.98}_{-1.11}$	$52.60^{+0.83}_{-0.93}$	$52.58^{+0.73}_{-0.63}$
12.00	$53.97^{+0.94}_{-1.24}$	$53.94^{+1.00}_{-1.45}$	$53.52^{+1.18}_{-1.27}$	$53.39^{+1.03}_{-1.10}$	$53.38^{+0.82}_{-0.84}$	$53.22^{+0.89}_{-0.76}$
12.50	$55.04^{+0.66}_{-0.99}$	$54.70^{+0.80}_{-1.25}$	$54.36^{+0.93}_{-1.79}$	$53.90^{+1.04}_{-1.18}$	$53.90^{+0.90}_{-0.90}$	$53.80^{+0.75}_{-0.88}$
13.00	$55.64^{+0.38}_{-1.15}$	$55.30^{+0.60}_{-0.97}$	$54.75^{+0.77}_{-1.57}$	$54.27^{+0.98}_{-1.56}$	$54.03^{+1.00}_{-1.12}$	$53.84^{+1.00}_{-0.93}$
13.50	$55.74^{+0.55}_{-1.06}$	$55.60^{+0.60}_{-0.93}$	$55.30^{+0.58}_{-1.23}$	$54.46^{+0.94}_{-1.59}$	$54.08^{+1.25}_{-1.38}$	$54.00^{+0.85}_{-1.03}$
14.00	$56.10^{+0.38}_{-0.85}$	$55.65^{+0.65}_{-0.76}$	$55.35^{+0.73}_{-1.21}$	$54.85^{+0.74}_{-1.44}$	$53.93^{+1.17}_{-1.35}$	$53.17^{+1.60}_{-0.67}$