SEMI-ANALYTIC FORECASTS FOR GALAXY FORMATION AND COSMIC REIONIZATION IN THE ULTRAHIGH-REDSHIFT UNIVERSE

By

LONG YAN AARON YUNG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

written under the direction of

Dr. Rachel Somerville

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2020

ABSTRACT OF THE DISSERTATION

Semi-Analytic Forecasts for Galaxy Formation and Cosmic Reionization in the Ultrahigh-Redshift Universe

By LONG YAN AARON YUNG

Dissertation Director:

Dr. Rachel Somerville

Concerning the early evolution history of our universe and distant galaxies that are expected to be uncovered with future telescopes, we present a collection of comprehensive, physically-backed *forecasts* for galaxy populations across cosmic time and their interaction with the cosmic environment. Our predictions cover galaxies spanning a remarkably wide range of mass and redshift, for which we provide a large variety of physical and observable properties, and from which we derive the production rate of energetic photons and the subsequent reionization history. These pre-eminent results are achieved with a custom-built modelling pipeline based on the versatile Santa Cruz semi-analytic model for galaxy formation and the new model components introduced in this work. This efficient modelling pipeline established an effective link between the 'ground-level' galaxy formation physics and the 'top-level' cosmological-scale observables, which also enables controlled experiments that help pinpoint the set of physical processes that have leading effects on cosmological events and quantifying their impacts. This work demonstrates the enormous potential of using semi-analytic models to inform the planning of optimal survey strategies, support scientific rationale for observing programs, and facilitate physical interpretation for observed objects. In anticipation of the soon-to-be-launched *James Webb Space Telescope*, a portion of our predictions are tailored to explore its capability on detecting faint, distant galaxies and on constraining key galaxy formation physics.

Acknowledgments

First and foremost, I would like to thank my thesis advisor, Rachel Somerville, for taking me under her wing and for being an extremely supportive and encouraging mentor. I thank her for patiently guiding me from the beginning and through times when I was confused and frustrated. I am grateful for everything she taught me and did for me throughout the journey. Rachel has been and will continue to be my role model as an astronomer, a collaborator, and a mentor.

I thank Steve Finkelstein, Gergö Popping, Romeel Davé, Harry Ferguson, Aparna Venkatesan, and Peter Behroozi for their tremendous contributions to the *Semi-analytic forecasts for JWST* paper series, which this dissertation work is based on. I also thank Andrew Baker, Saurabh Jha, Anirvan Sengupta, and Steve Finkelstein for serving on my thesis committee. The main chapters of this dissertation have been published in the Monthly Notices of the Royal Astronomical Society, which I also thank Assistant Editor Morgan Hollis for his assistance during the publication process.

I acknowledge the generous support from the Simons Foundation and the Downsbrough family, as well as funding from HST grant AR-13270.001-A from NASA, which jointly supported my Ph.D. study and travel expenses for presenting the results from this work at conferences and many research institutions.

I thank the Center for Computational Astrophysics at the Flatiron Institute for their warm hospitality throughout my time there as a guest researcher. I am grateful for the colleagues and staff members there and for the friendly, supportive, and inclusive work environment they have created. I also thank the many fellow astronomers I met at conferences, workshops, collaboration meetings, and visits, especially the ones who hosted me. If I were to list each and every one of you, the list will go on for pages. Instead, I am just going to thank all of you at once for spending time with me and for all the inspiring conversations we had.

The galaxy formation models that form the basis of this work were run primarily on computing cluster *Amarel* managed by the Office of Advanced Research Computing at Rutgers University. All post-processing and follow-up analysis were executed on computing cluster *rusty* managed by the Scientific Computing Core at the Flatiron Institute, who also manages the data products released alongside this work. I acknowledge the efforts and technical supports provided by the staffs from these places.

Last but not least, I thank my family and friends for all their support over the years. Thank you for bearing with me and being there for me when I need you. Thank you.

Dedication

To the marvelous creation of the almighty God and our curiosity.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

1.1 Cosmogonic Preliminaries

Despite being trapped on our own planet, our venture into the distant Universe is made possible by capturing photons coming from far-away objects. These photons traverse through both space and time, and are encoded with messages about the sources where they were originated and the conditions of the Universe along their paths. Perhaps the most interesting consequence that follows from the finite speed of light is that the farther a photon has travelled, the further back in time the message dates. This is sometimes referred to as the 'look-back effect'. For this reason, censusing the Universe at various distances allows us to piece together the evolution history of the objects within, and by extension, the Universe itself.

Another intriguing feature is the orderly colour shift observed for distant objects due to the Doppler effect, which occurs when a light source is in relative motion to the observer, in this case, ourselves. This effect shifts light to redder colours for objects travelling away from us (hence '*red-shifted*', denoted with symbol z), which the severity of the shift has been used to determine an object's recessional velocity along our line of sight. In 1929, Edward Hubble was the first to realize the correlation between redshift and distance for far-away objects, and their recessional velocities can be related to their distances by the Hubble constant (Hubble 1929). This discovery is now referred to as the Hubble's law, which has since been repeatedly tested and shown robust with advancing instrumentation and observation techniques (e.g. Riess et al. 2016; Planck Collaboration 2014). Because of such tight and strong correlation, redshift also doubles as the main diagnostics for the distances of observed objects.

Consider both the fact that all distant objects are redshifted (moving away from us) and the degree of shift is tightly correlated to their distances (farther objects travel faster), instead of arriving at the conclusion that we are sitting at the centre of the Universe, it is quite undeniable that the Universe we live in has been undergoing a centreless, accelerating expansion in every direction. Reversing the expansion in time, we can see that the past Universe was more compact and therefore hotter and denser. And winding the clock back even further, we will arrive at the point in time

where the entire Universe converges; it is referred to as the 'Big Bang', the point where space and time begin to exist. This epic point in time also left behind a few traces that help us validate and further understand how such startling prologue to the Universe unfolded.

The very first batch of baryons emerged under some extreme conditions, known as Big Bang nucleosynthesis (BBN). This process took place seconds to minutes after the Big Bang and is responsible for producing the most abundant elements in the Universe as observed today, including hydrogen, helium, and a trace amount of lithium (Alpher & Herman 1948; Peebles 1966; Wagoner et al. 1967). The specific proportion of the primordial baryonic composition depends on the condition of the Universe during the process, specifically the temperature and density. In some parts of the Universe where very few physical interactions occurred, the chemical composition remains *pristine* until today, which becomes a way to glimpse at the conditions of the primordial Universe. Furthermore, the broadly uniform chemical compositions and helium floor measured in nearby stars and galaxies across the Universe also point to the fact that their compositions are set by a cosmological scale event.

The hot Universe in the past also left behind thermal radiation that is still detectable today in the form of microwaves, known as the cosmic microwaves background (CMB). This signal coming from the edge of the visible Universe was first predicted by Alpher & Herman in 1948 and its first detection was made by Penzias & Wilson in 1965. This seemingly bizarre signal is found to be smooth and uniform, which indicates that it is not associated with specific celestial objects but rather coming from the Universe itself. This is also a strong evidence showing that the Universe is homogenous (same everywhere) and isotropic (same in every direction) on the largest scales. Over the years, multiple space missions were deployed dedicated to obtaining detailed, full-sky CMB maps (Smoot et al. 1992; Spergel et al. 2007; Planck Collaboration 2014). These maps are also embedded with crucial information that helps characterize our Universe, such as the overall matter density and the Hubble constant. The measured expansion rate of the Universe, the largely uniform CMB, and the primordial compositions of baryonic matter are direct evidence pointing to a hot beginning of the Universe, and together they form the three pillars of modern cosmology.

A cosmological model is an aggregate set of mathematical equations that prescribes the intrinsic characteristics of the Universe. The ΛCDM model, also referred to as the standard model of cosmology, is exemplary at reproducing many of the observed characteristics of our Universe (Frieman et al. 2008). Such a theoretical framework depicts a Universe mainly composed of a cosmological constant (Λ) and cold dark matter (CDM), where the former drives the universal expansion and the latter dominates the Universe's matter budget (Zwicky 1937; Perlmutter et al. 1999; Bertone et al. 2005). The standard model also implies that large-scale structures across the Universe were

formed in accordance with the hierarchical structure formation paradigm. That is, low-mass seeds first emerge from small-scale gravitational instability, and then grow by accreting mass from the cosmic environment and merging with objects nearby. These physical implications serve as the basis for galaxy formation models.

1.2 Galaxy Formation and Evolution

Though on large scales matter seems to be distributed across the Universe uniformly, looking closely at the sites that harbour structure formation, we find that matter tends to cluster in islands known as dark matter halos. Inside the gravitational well of these halos, baryonic matter further concentrates to form finer systems of stars, gas, and dust, assembling what we called galaxies. The abundance of galaxies and their physical properties are both important probes for the underlying physical processes and the conditions of the cosmic environment. Galaxies that are actively forming stars can be extremely luminous and therefore be detected from great distances. However, extracting information from observations is not at all straight forward given the wide variety of intricate physical processes that take part in these complex systems. For the same reason, building a galaxy formation model is highly non-trivial and remains one of the most active research areas in astrophysics

The formation of galaxies is a manifestation of many physical processes operating together. In the current picture of structure formation, galaxies form in the centre of the gravitational wells of large bodies of dark matter that are commonly referred to as dark matter halos. The formation of galaxies is a manifestation of many physical processes operating together. In the context of galaxy formation, physical processes broadly refer to the various kinds of interactions of baryonic matter that lead to energy transfer and phase transition. They can take effect locally (e.g. confined in small regions such as star clusters) or globally (e.g. across the entire galaxy), or both. Since these processes, for the most part, are decoupled from dark matter interactions, they are therefore also referred to as baryonic processes. One of the main goals of modern astrophysics is to identify the set of physical processes that take part in shaping galaxy formation. Some examples of key processes found in past studies include accretion and cooling of gas, conversion of cold gas into stars, production of metals and dust, and the fuelling and growth of supermassive black holes, etc. In addition, there are also special circumstances, such as galaxy mergers, that may cause starbursts, which are short periods of time when star formation is stimulated. This creates an ecosystem of processes that are vital to the production of radiation, including the essential LyC.

In some cases, the output of a process is returned back to the system and subsequently modifies the conditions that trigger the process. These *feedback effects* can lead to complicated and sometimes

counter-intuitive outcomes. For example SNe are responsible for producing and depositing metals back to the ISM, which become an important coolant that enables gas to cool more efficiently and yield a higher rate of star formation. However, the thermal and kinetic energy they release can also reheat and eject gas from galaxies, and in turns, suppress or even quench star formation. Other feedback effects will be described and discussed throughout this thesis. These effects can act simultaneously with other processes and give rise to even more complicated situations, and may be extended to affect nearby galaxies.

The collective effects of physical processes are reflected in the physical properties of galaxies. Some of the key properties that are frequently used to characterise the conditions of star forming regions in galaxies include star formation rate (SFR), stellar mass (M_*), stellar age, metallicity (Z), and the density of multiphase gas. Surveying and sorting the statistical distributions of properties in large galaxy populations provide an overview of the galaxy demographics at a given time in the Universe. And combining the distributions of two or more of these properties can reveal the scaling relations among them. Some examples of scaling relations are Faber-Jackson (luminosity and the velocity dispersion), Tully-Fisher (intrinsic luminosity and line width), Kennicutt-Schmidt (SFR and cold gas density), etc. These relations are historically important especially to the early study of galaxy formation physics.

The physical properties of astronomical sources are collectively exhibited in their spectral energy distribution (SED), which is the distribution of energy over wavelength or frequency. SEDs are a blend of radiation from stellar populations, AGN, and other emission and absorption features, and are the prime diagnostics for astronomical sources. In addition, the intervening IGM can also leave imprints on the SED. The most common scenario is that the redshifted spectra can be partially absorbed by neutral hydrogen. A noteworthy scenario is that the spectra of high-redshift objects can be dynamically absorbed by neutral hydrogen as it is being redshift, resulting in the Lyman-break feature in galaxies spectra and the Gunn-Peterson Trough in AGN spectra that can be used to constraint the onset of reionization. For distant objects, for which it is extremely hard to obtain their high-resolution spectra, their spectral properties are approximated with the integrated flux within some frequency bands (e.g. ultraviolet) and their optical colours. These are generally referred to as photometric surveys, and this technique is able to detect sources in a more time-efficient manner compared to spectroscopic surveys, at the expense of larger uncertainties on the redshifts of individual objects.

1.3 Reionization of Intergalactic Hydrogen

Outside of the overdense regions, the Universe is uniformly filled with diffuse gas; it is commonly referred to as the intergalactic medium (IGM). While the chemical composition remains pristine since the epoch of BBN, the temperature of the IGM gradually decreases as the Universe expands, allowing free electrons to bond with nuclei for the first time and resulting in a neutral IGM. As the temperature of the expanding Universe continues to drop, large-scale structures are able to form, which produces ionizing radiation that gradually (re)ionizes the IGM. It is evident that this global phase transition took place approximately 0.47 to 0.93 billion years after the Big Bang (or equivalently, $z \sim 10$ to 6), which is commonly referred to as the Epoch of Reionization (EoR). Although the process has not been fully understood, its onset and duration, as well as the intermediate stages, are jointly constrained by various kinds of observations. For instance, the presence of neutral hydrogen in the early Universe can partially absorb spectra of luminous quasars, imprinting unique spectral features that can be used to determine an upper limit for the time of the onset of the process. As patches of the ionized IGM grow in size over time, the gradual depletion of neutral hydrogen permits free passage for radiation, and therefore the Universe became transparent. On the other hand, the emerging free electrons along the line of slight can scatter and polarize the CMB. Therefore, the measured electron optical depth or polarization of the CMB has been used as an indicator for how long the Universe has been ionized.

The Lyman-continuum (LyC) radiation, which collectively refers to photons with energy $\gtrsim 13$eV, is responsible for the ionization of intergalactic hydrogen. However, both the overall ionizing photon budget and the characteristics of the sources are still under investigation. Recent theoretical and observational studies have come to some consensus that high-redshift galaxies are most probably the dominate sources of LyC during the EoR based on their projected high number density and the ability to produce LyC photons by the young stellar population therein. Although x-ray binaries and active galactic nuclei (AGN) are also known to produce ionizing photons during the EoR, their contribution to the overall ionizing photon budget are unlikely to be dominant due to their projected low abundance in the early Universe (Madau & Fragos 2017; Dayal et al. 2020; Trebitsch et al. 2020).

1.4 Interplay between the Big and the Small

The many physical processes that took part in galaxy formation, as described in the previous section, operate over a vast range of scales. This include spatial scales stretching from sub-parsec activities, such as individual stars forming and interacting within star clusters, to super-megaparsec galaxy mergers and formation of cosmic web structures. Similarly, the timescales over which these processes

operate can span from a few days (e.g. SNe reaching peak luminosity) to a fraction of the age of the Universe (e.g. our Milky Way Galaxy merging with the neighbouring Andromeda Galaxy). The effects of these physical processes can also propagate upward and downward across spatial and temporal scales, making the 'multi-scale' network of processes even more intricate. Furthermore, emerging galaxies are also affected by circumstances beyond the boundary of their host dark matter halos, such as the conditions of the cosmic environment (e.g. baryonic fraction and temperature of the IGM) and cosmological-scale events (e.g. the phase of the IGM during the EoR). These effects are especially pronounced at early times, affecting the first galaxies and the first stars. The conditions from these early stages will propagate through time and reflect in the present day the properties of galaxies and their associated observables. In turn, radiation and metals produced in galaxies can also escape from these halos and have extended effects on the cosmic environment and nearby objects.

Models and theoretical simulations set within the framework of cosmological structure formation are a powerful tool for understanding how the Universe operates and how galaxies formed. However, modelling galaxy formation in a cosmological context can be extremely challenging. As predicted by the hierarchical structure formation paradigm and backed by observations, massive halos are much more rare than their lower-mass counterparts. In order to simulate a halo population that meaningfully represents the ones in the real Universe, a simulation needs to be able to simultaneously include halos over a wide mass range, which requires an extremely large simulated volume to capture the rare, massive halos and a very high mass resolution to resolve the lower mass ones. Conventional numerical methods, such as hydrodynamic simulations and adaptive mesh refinement, simulate the formation of structure by tracking the migration of mass represented by a large number of 'particles' and 'gridcells'. The demand for computational resources by these methods increase extremely rapidly with simulated volumes and resolution. There are tradeoffs between the breadth of physical processes that can be included, the accuracy with which these processes can be modelled, and computational limitations. In addition, tracking the interactions of baryonic matter is even more challenging, where many of the baryonic processes operate in scales below the spatial resolution of these simulations. Simulating everything across different physical scales with conventional numerical method with needed resolution is very much impractical. Inevitably, simulations have to rely on 'sub-resolution' or 'sub-grid' recipes.

1.5 The Semi-Analytic Modelling Approach

Semi-analytic models (SAMs) provide an alternative, 'middle way' approach for modelling galaxy formation in comparison with conventional numerical methods. Unlike those methods, SAMs employ a set of carefully curated recipes to track the evolution of the global properties of galaxies. This includes standard components such as radiative cooling of gas, formation of stars, chemical evolution, feedback due to massive stars and supernovae, etc. It is worth noting that some of these recipes are very similar to the sub-grid recipes adopted in numerical simulations. These recipes are either derived analytically from first principle or empirically from observations or simulated results. The use of these recipes allows the SAMs to simultaneously track the effects of a wide variety of baryonic processes that operate across a vast range of scales and provide predictions for a wide range of physical properties, all at a relatively low computational cost. Given the great flexibility of the SAM and the high level of control we have over the model components, it is also a very handy tool for exploring how different assumptions may impact the outcomes.

The work presented in this dissertation is based on the foundation of the Santa Cruz semi-analytic models of galaxy formation that was first presented in the dissertation work by Somerville (1997). Since the debut of the modelling framework, it has undergone a few major modifications and improvements to better reflect the findings from latest observations and theory studies. Some examples are the recently implemented recipes for multiphase gas partitioning, based on numerical simulations, and an H_2-based star formation recipe, based on nearby galaxy observations (Popping et al. 2014; Somerville et al. 2015). It has been shown to be one of the most comprehensive, well-established galaxy formation models currently available.

The technical details of the modelling framework for galaxy formation used in this work are described in full in Section 2.2. Additional model components are described in the respective sections where relevant (Section 3.2, 4.2, and 5.2.1). It is worth noting that both semi-analytic and numerical simulation approaches required the use of 'tunable' parameters that are need to be calibrated using observations. The calibration of the free parameters is given in Appendix B.

1.5.1 Merger Tree Construction

The SAM relies on prior knowledge of the merger histories of dark matter halos, commonly referred to as 'merger trees', which can either be extracted from numerical simulations or constructed using a semi-analytic approach. Based on our understanding of gravitation and how dark matter halos emerge and collapse into bound, virialized structures in a ΛCDM Universe, it is possible to summarize these behaviours in an empirical, statistical sense with a set of mathematical formalism. The

Extended Press-Schechter formalism lays down the fundamental statistical prescriptions for structure formation and growth in a hierarchical Universe (Press & Schechter 1974; Bond et al. 1991; Bower 1991; Sheth et al. 2001; Sheth & Tormen 2002), which provides the necessary mathematical framework for the construction of dark matter halo merger trees (Lacey & Cole 1993, 1994; Somerville & Primack 1999; Cole et al. 2000; Parkinson et al. 2008; Zhang et al. 2008). These algorithms are able to create a large number of Monte Carlo realizations of merger trees, while requiring a small fraction of computational resources needed for a typical cosmological-scale hydrodynamic simulations.

The merger tree algorithm used in this work is based on the method presented in Somerville & Kolatt (1999) with modification documented in Somerville et al. (2008). We note that developing such an algorithm is highly non-trivial as a model that seems to perform well at low redshifts does not guarantee good performance at higher redshifts. Among the many methods explored in Somerville & Kolatt (1999), the N-Branch Tree Method with accretion had been shown to be the most successful, practical solution. Here we provide a concise summary of the algorithm.

For a given halo with mass M_0 at redshift z_0, also referred to as the 'root halo', the masses of its 'progenitor halos', M, at an earlier time $z > z_0$ can be drawn from the mass-weighted probability distribution (reproduced from Lacey & Cole (1993) in the notation of Somerville & Kolatt (1999))

$$P(\Delta S, \Delta\omega) d\Delta S = \frac{1}{\sqrt{2\pi}} \frac{\Delta\omega}{(\Delta S)^{3/2}} \exp\left[-\frac{(\Delta\omega)^2}{2\Delta S}\right] d\Delta S, \tag{1.1}$$

where $S \equiv \sigma^2(M)$ is the mass variable and $\omega \equiv \delta_c(t)$ is the time variable. These quantities are defined to work naturally with a smoothed Gaussian density field that grows linearly over time. This conditional probability effectively represents the distribution of progenitor number and mass dictated by a given cosmological matter density and mass variance. By iterating the process of drawing progenitors from the distribution, one can construct a probable trajectory for a halo's past merger history.

For a predefined resolution mass $M_{\rm res}$, if the randomly drawn $M > M_{\rm res}$, it is regarded as a progenitor. Otherwise, it is regarded as accreted mass. This procedure is repeated for the unallocated mass $\Delta M = M_0 - M$ until the remainder reaches $\Delta M < M_{\rm res}$. This method allows two or more progenitors to be identified depending on the mass of the progenitor (thus N-Branch). Each identified progenitor will then undergo the same procedure to have their progenitors identified. This process ends when the remainder mass is below $M_{\rm res}$. The time-step for the process is defined by (reproduced from Somerville & Kolatt (1999))

$$\Delta\omega \lesssim \sqrt{\frac{dS}{dM}(M_0)\Delta M_c}\ , \tag{1.2}$$

which is chosen carefully to avoid the breakdown of the association of halos across two generations, which occurs when the time-step is too big, or the domination by numerical noise when the step is too small.

The outcomes of this algorithm are checked against the progenitor mass function, the number-mass distribution of progenitors, and evolution of halo mass functions across redshifts predicted by the Press-Schechter theory. It has been shown that these constructed trees are in statistical agreement with results from N-body simulations. For the efficiency and flexibility of this merger tree construction approach, we are able to cover an extremely wide range of mass and redshift in our prediction, with a mass resolution much higher than the ones obtained from numerical simulations.

1.5.2 Model Extension for Cosmic Reionization

Some of the principles and techniques from modelling galaxies can be modified and adopted to approximate the condition of the multiphase IGM. In this work, we implemented an analytic IGM model (see Section 5.2.2 for full descriptions) to track the overall reionization history of the intergalactic hydrogen based on the ionizing photon budget supplied by the predicted galaxy populations. By doing so, we create an extended modelling pipeline that effectively bridges the gap between galaxy formation physics and large-scale reionization history. This allows us to assess whether these predicted high-redshift galaxy populations are capable of reionizing the Universe in the time frame required by observations.

However, given that this modelling method handles most galaxy components as bulk quantities, it faces limitations in predicting morphological and geometrical features. Therefore, we have to make simple assumptions for the internal structure and the spatial distribution for matter. Similarly, the pipeline is only able to predict the global fraction of ionizing hydrogen but is not able to track the geometry of reionization nor represent the progress in regions that deviate much from the mean matter density (e.g. near a galaxy cluster or in a void).

1.6 Observations of the Early Universe

Direct detections of extremely distant objects serve as the basis for the study of galaxy evolution and early history of our Universe. Over the past decades, significant efforts have been dedicated to advancing instrumentation technologies and observing techniques, which made the construction of the acclaimed space-based *Hubble Space Telescope* (*HST*) possible. Since its deployment, *HST* has been constantly pushing the frontier of deep space exploration. In 1995, *HST* carried out the monumental *Hubble* Deep Field (HDF) survey, revealing a distant Universe full of galaxies that had

never been seen before (Williams et al. 1996). In the following decades of service, *HST* continues to deliver remarkable breakthroughs, reaching objects even fainter and farther with the *Hubble* Ultra Deep Field (HUDF; Beckwith et al. 2006) and eXtreme Deep Field (XDF; Illingworth et al. 2013). In addition, *HST* also plays an important role in large survey programs, such as the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), which typically covers significantly larger area than HUDF-like programs. In special circumstances, a survey can reach even fainter objects in fields that are gravitationally lensed by massive foreground clusters (Lotz et al. 2017), as the gravitational field around these massive clusters can effectively act as a magnifying lens by bending light from objects behind them.

Overall, *HST* together with ground-based observatories have found tens of thousands of galaxies, about 2,000 of which are within the first 10% of the Universe's history. These observations put constraints on the number counts of objects and their UV luminosities, which can be used to infer a number of physical properties, including star formation rate and stellar mass. However, the detection of distant objects are currently limited to the relatively luminous objects at their time due to instrument limitations, leaving the bulk majority of low-mass objects unconstrained. These high-redshift, low-mass galaxies are thought to have played significant roles throughout cosmic history and could also be analogous to the progenitors of our own Milky Way Galaxy. Nevertheless, current theories for galaxy formation predict a wide range of possible scenarios with these undetected galaxy populations. Uncovering these galaxy populations, as well as constraining the physical processes that operate at these epochs, will rely on future generations of instruments.

1.7 Future Explorations with the James Webb Space Telescope

The soon-to-be-launched *James Webb Space Telescope* (*JWST*), successor to the *HST*, is a large, cryogenic, infrared space-based observatory (Gardner et al. 2006). It will be situated 1.5 million kilometres away from the Earth in the second Lagrangian (L2) point, which is a position that remains nearly stable relative to the Earth-Sun system while orbiting around the Sun. By keeping it in the shadow cast by the Earth, the multilayer sunshield and cryocooler are able keep the extremely sensitive on-board instruments to their functional temperatures (a few to a few tens of K). *JWST* is also equipped with a primary mirror 6.5 meters in diameter, the largest of all space telescopes, that gives it unprecedented light-gathering power. These design and technological advancements together achieve sensitivity that is ~ 100 times that of *HST*, empowering *JWST* to detect extremely distant and faint galaxies.

One of the main scientific objectives of *JWST* is to constrain the nature of the sources that

reionized the Universe. *JWST* will be able to efficiently census faint sources with the photometric instrument Near Infrared Camera (NIRCam). The Mid-Infrared Instrument (MIRI) and Near Infrared Spectrograph (NIRSpec) will be able to conduct follow-up spectroscopic diagnosis for detected candidates. Together these instruments enable both photometric and spectroscopic surveys for galaxies over a wider mass and luminosity range, which also allows for the first time direct detections for faint objects in the early Universe. *JWST* is expected to deliver high quality data that will lead to breakthroughs in our understanding of the very early Universe.

In this work, we use the published *JWST* instrument specifications as an example to relate predictions made with theoretical models to what can practically be done with observations. We also use these results to assess the capabilities and make forecasts for this long-anticipated flagship telescope, especially for deep-space photometric surveys. The predictions and tools provided in this work can be used to back the scientific rationale of proposed observing programs, help optimize the observing strategies for upcoming surveys, and facilitate physical interpretation for future observations. The predictions from the physical model are independent of any instruments and can be easily adapted to represent other telescopes or observing programs in the future.

1.8 The Contents of this Thesis

In this dissertation, we present a wide variety of predictions for the high-redshift galaxy populations and related cosmological observables that are expected to be detected by *JWST* or other future instruments. In Chapter 2, we present distribution functions for the rest-frame UV luminosity and observed-frame IR magnitudes in the full set of *JWST* NIRCam broadband filters. In Chapter 3, we further investigate the physical properties and the scaling relations for galaxies predicted by the same models. In Chapter 4, we implement and examine a new and crucial model component that makes predictions for the intrinsic production rate of ionizing photons by star-forming galaxies based on the physical properties of the predicted stellar populations. In Chapter 5, we combine our galaxy formation model with an analytic reionization model and a parametrized treatment of the escape fraction to create a physically motivated, source-driven pipeline to efficiently explore the effects of the predicted galaxy populations on cosmic reionization. In Chapter 6, we provide a concise summary of our results. All results presented in this work, including tabulated data and full object catalogues, are made available at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/.

Chapter 2

UV Luminosity Functions of High-Redshift Galaxies

In anticipation of the upcoming deployment of the *James Webb Space Telescope* (*JWST*), we present high-redshift predictions by the well-established Santa Cruz semi-analytic model. We update the models by re-calibrating them after adopting cosmological parameters consistent with recent constraints from Planck. We provide predictions for rest-frame UV luminosity functions for galaxy populations over a wide range of $M_{\rm UV}$ from ~ -6 to ~ -24 between $z = 4$–10. In addition, we present the corresponding predictions for observed-frame galaxy number counts in different redshift bins in the full set of NIRCam filters. We provide predictions of the quantitative effect on these observables of varying the physical recipes implemented in the models, such as the molecular gas depletion time (star formation efficiency) scalings or the scalings of outflow rates driven by stars and supernovae with galaxy circular velocity. Based on these results, we discuss what may be learned about the physical processes that shape galaxy formation from *JWST* observations of galaxy number densities at different intrinsic luminosities.

This chapter is reproduced from published work *Semi-analytic forecasts for JWST – I. UV luminosity functions at* $z = 4$–10 (Yung et al. 2019. MNRAS, 483, 2983).

2.1 Introduction

The soon-to-be-launched *James Webb Space Telescope* (*JWST*) will possess the unprecedented infrared sensitivity and spatial resolution required for detecting faint, distant galaxies that are extremely difficult or impossible to detect with any current facilities. These observations will provide significant insights into the statistical properties of the galaxy population near cosmic dawn. This is of great interest for determining whether the physical processes that shape galaxy formation are very different in the early universe from locally, as well as for constraining which objects are responsible for reionizing the Universe.

Hubble Space Telescope (*HST*) has detected nearly 2000 galaxy candidates at high redshifts ($z \sim 4$–10) from both blank and gravitationally lensed fields (Koekemoer et al. 2013; Lotz et al. 2017). Additionally, brighter objects have also been discovered with ground-based facilities, such as

the United Kingdom Infra-Red Telescope (UKIRT) and the Visible and Infrared Survey Telescope for Astronomy (VISTA) (McLure et al. 2009; Bowler et al. 2015). These observations have provided constraints on the space density of relatively bright galaxy populations up to $z \sim 10$ (e.g. McLure et al. 2009, 2013; Castellano et al. 2010; van der Burg et al. 2010; Oesch et al. 2013, 2014, 2018; Schenker et al. 2013; Tilvi et al. 2013; Bowler et al. 2014, 2015; Bouwens et al. 2014a, 2015b, 2016b, 2017; Schmidt et al. 2014; Atek et al. 2015; McLeod et al. 2015, 2016; Finkelstein et al. 2015a; Livermore et al. 2017; Ishigaki et al. 2018). However, the constraints on the faintest populations currently rely solely on fields that are lensed by massive foreground galaxy clusters, and the corrections for magnification are quite uncertain (e.g. Kawamata et al. 2016; Bouwens et al. 2017; Priewe et al. 2017). *JWST* will probe much deeper down the luminosity function in unlensed fields, as well as providing more secure redshift measurements for high-z candidates that are currently selected by the Lyman-break technique. Many members of the community are currently engaged in planning the optimal strategies to take advantage of the limited lifetime of *JWST* to achieve these breakthroughs in our understanding of the very early Universe. Theoretical predictions may be able to aid in designing the observing strategy and trade-offs in imaging area, depth, and wavelength coverage. Perhaps more importantly, it is critical to assess the *uncertainties* on current theoretical predictions and the plausible range in galaxy properties at these extreme redshifts.

With the cosmological parameters that govern early structure formation fairly well constrained by experiments such as the Wilkinson Microwave Anisotropy Probe (WMAP; Spergel et al. 2003, 2007; Komatsu et al. 2009, 2011; Hinshaw et al. 2013) and Planck satellite (Planck Collaboration 2014, 2016a), combined with other constraints from Baryon Acoustic Oscillations, Supernovae, and weak lensing (March et al. 2011; Aubourg et al. 2015; Hildebrandt et al. 2017), the major uncertainties in forecasting the abundance of galaxies in the early Universe arise from our lack of a rigorous theory for how dense, cold molecular clouds form, how stars form within these clouds, and how thermal energy, momentum, and radiation from stars and supernovae affect the subsequent efficiency of star formation ('stellar feedback'; see McKee & Ostriker (2007), Somerville & Davé (2015, hereafter SD15), and Naab & Ostriker (2017, hereafter NO17)). Although feedback from radiation and jets produced by accretion onto supermassive black holes is likely to be critical for regulating galaxy formation in the lower redshift Universe (SD15 and references therein), at very early times $z \gtrsim 6$ most galaxies probably have not had time to form massive black holes in their nuclei, so it is likely (although not certain), that this form of feedback is sub-dominant at these epochs.

Some theorists have suggested that star formation remains very efficient out to very high redshifts (e.g. Behroozi & Silk 2015), while others have suggested that the low metallicity environments will make it difficult to form molecular hydrogen and lead to inefficient star formation (e.g. Krumholz

& Dekel 2012). The efficiency of star formation in low-mass halos at early times is also critical to determining which objects reionized the Universe, and how reionization proceeded in space and time. Numerous studies have shown that the already detected galaxy population is likely insufficient to reionize the Universe by $z \sim 6$–10 as required by observations (Finkelstein et al. 2015a; Robertson et al. 2015), and it is generally assumed that the shortfall in the ionizing photon budget is made up by faint galaxies that form in low-mass halos (Bouwens et al. 2012; Kuhlen & Faucher-Giguère 2012; Atek et al. 2015; Gnedin 2016; Anderson et al. 2017). Many published theoretical predictions for reionization assume a fixed efficiency for converting baryons into stars in halos (Vogelsberger et al. 2013; Hopkins et al. 2014; Mutch et al. 2016). However, it is known that in the very nearby Universe, star formation is extremely inefficient in low-mass halos, leading to questions about whether the local and high redshift observations are in tension.

Computational methods have long been used in modelling the formation and evolution of galaxies. However, given that these processes operate over a vast range of scales, both spatially and temporally, modelling galaxy formation in a cosmological context remains one of the greatest challenges in astrophysics today (SD15). One possible approach to predicting galaxy properties at high redshift is to constrain the relationship between galaxy and dark matter halo properties at lower redshift, and then assume that these relationships hold out to higher redshift (e.g. Behroozi & Silk 2015; Mason et al. 2015; Furlanetto et al. 2017). Another is to implement physical recipes for the key processes that are thought to shape galaxy formation (e.g. gas accretion and cooling, star formation and stellar feedback, chemical enrichment, black hole formation and feedback) within a cosmological framework (see SD15, NO17 for a recent review). These physical recipes may be implemented within numerical simulations, which attempt to explicitly follow the equations of gravity, hydrodynamics, and thermodynamics for particles or grid cells (Springel et al. 2001; Bryan & Norman 1997, 1999). However, current cosmological simulations are unable to directly resolve the multiphase structure of the interstellar medium, the formation and evolution of individual stars and supermassive black hole (SMBH), or their interaction with their surroundings. Thus, 'sub-grid' recipes must be used to represent these processes on scales below those that can be explicitly resolved in the simulations. In most current numerical simulations, these sub-grid recipes are generally phenomenological and contain free parameters that must be calibrated to match global galaxy observations (see discussion in SD15). Due to the high computational expense, rather limited dynamic range can be achieved with these techniques, and the ability to explore different sub-grid recipes and parameter values is also limited.

An alternative is semi-analytic simulations, which apply simplified recipes for these same physical

processes within dark-matter halo 'merger trees' either extracted from dissipationless N-body simulations or created using semi-analytic Monte Carlo techniques. Despite the absence of a spatially defined 'grid' in semi-analytic models (SAMs), the essence of using phenomenological or physically driven recipes to model galaxies is very similar to the motivation of sub-grid recipes in conventional numerical simulations. One way to view SAMs is as, essentially, a book-keeping scheme that tracks the movement of mass between different reservoirs, as gas accretes from the diffuse intergalactic medium (IGM) into galactic halos, flows from halos into the interstellar medium (ISM) of galaxies, is converted from ISM into stars, and potentially is ejected from the ISM back into the hot diffuse galactic halo or back out into the IGM. In addition to gas and stars, heavy elements can also be tracked as they are produced by stars and circulate through the cosmic baryon cycle. Like the numerical simulations, the physical recipes in SAMs are phenomenological, and contain free parameters that must be calibrated to global galaxy observations.

This method has been widely used to explore a very broad range of galaxy properties (White & Frenk 1991; Kauffmann et al. 1993; Cole et al. 1994; Somerville & Primack 1999; Croton et al. 2006; Mutch et al. 2016) and it has shown that the predictions for the statistical properties of galaxies obtained using these methods are in very good agreement with those from numerical simulations (SD15). Due to the much greater computational efficiency and flexibility of these methods, however, they provide several advantages over the numerical approach, namely by allowing researchers to explore a broader range of parameter space and different physical recipes. Moreover, SAMs are also able to span a larger dynamical range in dark matter halo mass, from the smallest halos that are believed to be able to form stars to the most massive and rare objects.

Recently, many other groups have made theoretical predictions for high-z *JWST* observations using numerical (e.g. Barrow et al. 2017), phenomenological(e.g. Williams et al. 2018), and semi-analytic (e.g. Cowley et al. 2018) methods. These works collectively probe a broad range of aspects, including the physical and photometric properties of high-z galaxies (Cowley et al. 2018), population synthesis and synthetic spectra (Barrow et al. 2017; Volonteri et al. 2017), and the impact of broad-band filter choice on estimating photometric redshifts and on recovering their physical properties from observations (Bisigello et al. 2016, 2017). Despite the great interest in using *JWST* to uncover the physical nature and the assembly histories of high-z galaxies, however, only a handful of studies have systematically varied the underlying physics or model parameters in a controlled way within the same modelling framework. We find that these kinds of predictions will be quite informative for the interpretation of *JWST* observations.

In this chapter, we use the well-established Santa Cruz SAM, which has been shown to successfully reproduce key galaxy observations at lower redshifts ($z \lesssim 6$), to make predictions for galaxy

populations at intermediate to high redshifts ($z = 4$–10). Importantly, we make use of the Santa Cruz SAM version that tracks multiple phases of gas in the ISM (atomic, molecular, and ionized) using a suite of recipes based on empirical considerations or predictions from detailed numerical simulations containing treatments of molecular chemistry and radiative transfer (Popping et al. 2014, hereafter PST14; Somerville et al. 2015, hereafter SPT15). Our fiducial model is based on the empirically grounded assumptions that stars form in environments that are dominated by molecular gas, and that the formation of molecular gas depends on the gas surface density, metallicity, and the local UV radiation field. However, we also explore how sensitive our predictions at ultra-high redshift are to our recipes for star formation and stellar feedback. In this chapter, we focus on first-order, directly observable quantities, namely one-point distribution functions for rest-frame UV luminosity and observed-frame magnitudes in *JWST* NIRCam filters. In the rest of this thesis, we will explore physical properties of galaxies such as stellar masses, molecular gas content, metallicities, and dark matter halo masses, and will also make direct predictions for the implications for reionization.

The key components of this chapter are summarized as follows: the basic elements of the Santa Cruz SAM used in this work are summarized in Section 2.2. We then present the rest-frame UV LFs in Section 2.3 and counts in observed *JWST* filters in Section 2.4. We discuss our results in Section 2.5, and summarize and conclude in Section 2.6.

2.2 The Semi-Analytic Framework

The SAM used in this study is the same one outlined in Somerville et al. (2015, hereafter SPT15). Hence, we will not describe the model in full here and we refer the reader to the following works for full detail for the modelling framework developed by the Santa Cruz group: Somerville & Primack (1999); Somerville, Primack, & Faber (2001); Somerville et al. (2008, 2012); PST14 and SPT15. Throughout this work, we adopt cosmological parameters that are consistent with the ones reported by Planck Collaboration in 2015: $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8\,{\rm km\,s^{-1}Mpc^{-1}}$, $\sigma_8 = 0.831$, and $n_s = 0.9665$.

Dark matter halo merger histories, more commonly known as 'merger trees', are the backbone of the semi-analytic modelling framework. These merger trees can either be extracted from dissipationless N-body simulations or constructed using semi-analytic methods based on the Extended Press-Schechter (EPS) formalism (Press & Schechter 1974; Lacey & Cole 1993). In this work, in order to maximize the dynamic range and computational efficiency, we adopt the EPS-based method of Somerville & Kolatt (1999) with updates as described in Somerville et al. (2008, hereafter S08). We have also run our models using halo merger trees from the Bolshoi Planck N-body simulations,

and find very similar results over the dynamical range spanned by that simulation. However, in addition to having inadequate mass resolution to resolve the low-mass halos that host the faint galaxy population that will be detectable by *JWST*, Bolshoi Planck has only 29 snapshots stored above $z = 6$, which is inadequate to construct accurate merger histories for $z > 6$ galaxies. In addition, robust identification of halos at high redshift becomes tricky. Friends-of-friends methods, such as used in e.g. the BlueTides simulation (Feng et al. 2015) and the DRAGONS simulation suite (Poole et al. 2016), have been shown to artificially link together distinct halos at high redshift (Klypin et al. 2011). Although we are in the process of generating a new suite of simulations, halo catalogs, and merger trees that are carefully designed to model faint galaxies in the high-redshift Universe (Yung et al. in prep), we are confident that the EPS-based approach has sufficient accuracy for the somewhat qualitative exploration that we present here.

At each output redshift, we set up a grid of one hundred root halos with masses spanning the range in virial velocity $V_{\rm vir} \approx 20 - 500\,{\rm km\,s^{-1}}$, which covers halos ranging from close to the atomic cooling limit to the rarest objects expected to be detected in high-redshift surveys. We then weight each of these root halos by the expected abundance of dark matter halos of the given mass at the respective redshift, using the fitting functions provided in Rodríguez-Puebla et al. (2016) based on results from the MultiDark suite (Klypin et al. 2016) of N-body simulations. Although this involves extrapolating the fitting functions to lower halo masses and higher redshifts than those that are directly probed by the MultiDark Suite, we have validated these results using an unpublished suite of very high resolution, small box simulations kindly made available to us by Eli Visbal (Visbal et al. 2018). The assembly history is then traced down to a minimum progenitor mass of $M_{\rm res}$, which we refer to as the mass resolution of our simulations; here we set $M_{\rm res} = 10^{10}{\rm M}_\odot$ or 1/100th of the root halo mass, whichever is smaller. For each root halo in the grid, one hundred Monte Carlo realizations of the merger histories are generated.

After constructing these semi-analytic merger trees, our SAM implements a suite of fairly standard recipes such as cosmological accretion and cooling, star formation and stellar-driven winds, chemical evolution, black hole feedback, and mergers. We again refer readers to S08 and SPT15 for full details.

2.2.1 Gas Partitioning and Star Formation

In the most recent iteration of the Santa Cruz SAM (PST14, SPT15), discs are subdivided into annuli and the cold gas in each annulus is partitioned into an atomic (HI), ionized (HII), and molecular (H_2) component. In PST14 and SPT15, several different recipes for gas partitioning are investigated, including an empirical recipe in which the molecular fraction is determined by the disc

mid-plane pressure, and several variants of recipes in which the molecular fraction is determined by the gas surface density, metallicity, and the intensity of the local UV radiation field. Somewhat surprisingly, these studies found that most results were not very sensitive to which gas partitioning recipe was used. However, they did find that the recipe based on the prescription of Krumholz et al. (2009), in which the dependence on the UV background was not taken into account, failed to reproduce sufficient numbers of low-mass galaxies and was disfavored. Overall, the metallicity and UV-background dependent recipe based on simulations by Gnedin & Kravtsov (2011, hereafter GK) was found to perform the best, and was adopted as the 'fiducial' model. As we are concerned that some assumptions contained in the empirical pressure-based model may no longer hold at high redshift, we do not explore it here.

The usual picture is that the first stars form out of 'primordial' molecular hydrogen, form Pop III stars, and pollute early halos with metals (Frebel et al. 2009; Wise et al. 2012). In the current work we do not model the formation of metal-free Population III stars explicitly. Instead (as in previous works), we set a metallicity floor of $Z_{\text{pre-enrich}}$ to represent pre-enrichment from Pop III stars. We adopt $Z_{\text{pre-enrich}} = 10^{-3} Z_{\odot}$ (Bromm et al. 1999). In SPT15, we show that our results do not depend sensitively on the choice of this value.

Many earlier generations of SAMs adopt what we refer to as the 'classic' Kennicutt-Schmidt (KS) relation to model the rate at which cold gas is converted into stars. In this approach, the star formation rate (SFR) is assumed to scale as a power of the total (cold; HI+H_2) gas density (Schmidt 1959, 1963; Kennicutt 1989, 1998), where in some implementations (as in S08), only gas above a fixed surface density is assumed to participate in star formation. This may approximate to first order the transition from predominantly molecular to predominantly atomic gas that the above multiphase modelling attempts to capture, but there is observational evidence that this critical surface density depends on gas metallicity (Wilson 1995; Arimoto et al. 1996; Bolatto et al. 2008, 2011; Genzel et al. 2010; Leroy et al. 2011).

A more recent generation of SAMs adopts a H_2-based star formation recipe (Lagos et al. 2011; Fu et al. 2012, 2013; Somerville et al. 2015; Xie et al. 2017). Observations of nearby spirals have shown that the SFR surface density is nearly linearly proportional to the molecular hydrogen surface density (Wong & Blitz 2002; Bigiel et al. 2008, 2011; Leroy et al. 2011). However, these observations only probe H_2 surface densities up to about 50-80 $M_{\odot}pc^{-2}$. Mounting evidence from both observation and theory suggests that the slope of the SF relation may steepen to ~ 2 at higher surface densities (Sharon et al. 2013; Rawle et al. 2014; Hodge et al. 2015; Tacconi et al. 2018). We have explored both a 'single slope' SF relation (which we refer to as Big1), in which the molecular gas depletion time is effectively invariant with both galaxy properties and redshift, and a 'two slope' relation (referred to

as Big2), in which the molecular gas depletion time decreases (and star formation efficiency increases) with increasing H_2 surface density. The surface density of SFR is given by the expression:

$$\Sigma_{\rm SFR} = \frac{A_{\rm SF}}{\tau_{*,0}} \left(\frac{\Sigma_{\rm H_2}}{10 {\rm M}_\odot {\rm pc}^{-2}} \right) \left(1 + \frac{\Sigma_{\rm H_2}}{\Sigma_{\rm H_2,crit}} \right)^{N_{\rm SF}}, \tag{2.1}$$

where the critical H_2 surface density $\Sigma_{\rm H_2,crit} = 70$ ${\rm M}_\odot {\rm pc}^{-2}$, $A_{\rm SF}$ is the SF relation normalisation, and $\tau_{*,0}$ is a tunable normalisation parameter. We adopt a value of $A_{\rm SF}$ from observational determinations of the relevant Kennicutt-Schmidt relation (as given in Table 2.1), and allow $\tau_{*,0}$ to vary by about a factor of 50% up or down, reflecting the observational uncertainty in the true normalisation of the Kennicutt-Schmidt relation (although here we find that $\tau_{*,0} = 1$ produces good results). In SPT15, we found that although at $z \sim 0$, the predictions for galaxy properties were not very sensitive to which SF relation was adopted, an increasingly large discrepancy was seen at higher redshifts, up to the highest redshifts $z \sim 6$ explored in that paper. Accordingly, in this work we explore the implications of adopting these different SF recipes at even higher redshifts. The model variants are summarized in Table 2.1.

Table 2.1 Summary of the gas partitioning (GP) and star formation (SF) model variants explored in this work, where $N_{\rm SF}$ is the SF relation slope. GKBig2 is our fiducial model.

Model	GP recipe	SF law	$N_{\rm SF}$	$A_{\rm SF}$
KS	None	KS	1.4	1.1×10^{-4}
GKBig1	GK	Big1	1.0	4.0×10^{-3}
GKBig2 (Fid.)	GK	Big2	2.0	4.0×10^{-3}

2.2.2 Photoionization Feedback

Gas accretion and, ultimately, star formation activity in a galaxy can be reduced in the presence of a strong photoionizing background (Efstathiou 1992; Thoul & Weinberg 1996; Quinn et al. 1996); this effect is sometimes known as photoionization 'squelching' (Somerville 2002). Studies have shown that this effect is especially effective in low-mass halos, and it is thought to be significant in suppressing the collapse of gas into small mass halos and preventing the overproduction of dwarf galaxies in the local universe (Bullock et al. 2000; Benson et al. 2002b,a; Somerville 2002).

As in our previously published models, we adopt the approach proposed by Gnedin (2000) to model photoionization squelching, in which the fraction of baryons that can collapse into halos of a given mass $M_{\rm halo}$ at redshift z in the presence of a photoionizing background is computed using the

following function:

$$f_{\rm b}(M_{\rm halo}, z) = \langle f_{\rm b} \rangle \left\{ 1 + (2^{\alpha/3} - 1) \left[\frac{M_{\rm halo}}{M_{\rm char}(z)} \right]^{-\alpha} \right\}^{-3/\alpha}, \tag{2.2}$$

where $\langle f_{\rm b} \rangle$ is the cosmic mean baryon fraction, and $M_{\rm char}(z)$ is the mass at which halos retain half of the universal baryon fraction. In the present work, we adopt the redshift dependent characteristic mass obtained by Okamoto et al. (2008) from hydrodynamic simulations including a uniform metagalactic background (based on tabulated results kindly provided to us by T. Okamoto), and adopt $\alpha = 2$ as favored by those authors. We switch squelching on at a fixed redshift $z_{\rm squelch} = 8$, which is consistent with the redshift at which instantaneous reionization occurred as reported by the Planck Collaboration (2016b).

The characteristic mass $M_{\rm char}$ that we adopt ranges from a value of 1.4×10^7 $\rm M_\odot$ just after reionization to a value of $\sim 9.3 \times 10^9$ $\rm M_\odot$ (corresponding to a virial velocity of ~ 25 km/s) at $z = 0$. Note that these currently favored values of the characteristic mass (Hoeft et al. 2006; Okamoto et al. 2008) are significantly smaller than the values obtained by Gnedin (2000), and used in previous versions of the Santa Cruz SAM.

Our treatment of squelching is rather crude and of course is not self-consistent. Some previous semi-analytic studies have attempted to model photoionization squelching in a more self-consistent manner, based on the predicted UV emissivity of the galaxy population that emerges in the SAM (Benson et al. 2002b,a), and some numerical simulations have also included photoionization squelching self-consistently (Finlator et al. 2012; Gnedin & Kaurov 2014). However, these studies have found that the self-consistent modelling of squelching compared with the simpler approach adopted here does not have a significant effect on predictions for observable galaxy properties or reionization (Mutch et al. 2016). Indeed, as we show in Fig. 2.3, the Okamoto et al. filtering mass is so low that squelching has an almost undetectable effect on any observable galaxy properties at the redshifts that we investigate here. This is in agreement with the result found by Kim et al. (2013).

2.2.3 Black Hole Growth and Feedback

Top-level halos are seeded with black holes with an initial mass of $10^4 \rm M_\odot$. Black hole accretion and growth may occur in two modes. Mergers and disc instabilities trigger relatively rapid, radiatively efficient accretion (see S08 and Hirschmann et al. (2012) for details). In addition, a less rapid, radiatively inefficient mode of fueling is associated with Bondi accretion from the hot halo (again, see S08 for details). Two modes of AGN feedback are also implemented, associated with each accretion mode. The radiatively efficient mode drives winds that eject material from the cold gas

reservoir and drive it out of the halo. The radiatively inefficient mode is associated with relativistic jets that are assumed to heat the hot halo gas, reducing or even quenching cooling. The two models of AGN feedback have associated parameters, $\epsilon_{\rm wind}$ (see Eqn. 17 of S08) and $\kappa_{\rm AGN}$ (equivalent to $\kappa_{\rm radio}$ in Eqn. 20 of S08). The parameter $\epsilon_{\rm wind} = 0.5$ is set based on results from hydrodynamic simulations of binary galaxy mergers (see S08). The value of the parameter controlling the 'jet mode' feedback ($\kappa_{\rm AGN}$) is adjusted to fit constraints on the relationship between stellar mass and halo mass from abundance matching, and the stellar mass function derived from observations (see Appendix B).

In practice, we find that the implementation of AGN feedback currently included in our model has no noticeable impact on the predictions presented in this chapter – our results remain unchanged when we switch off both modes of AGN feedback. However, AGN feedback does impact the calibration of the model shown in the Appendix, and so we describe it here for completeness.

2.2.4 Stellar Feedback

Star formation and the baryon fraction in galaxies is thought to be regulated by large-scale outflows driven by massive stars and supernova explosions. The details of how these winds are driven and what determines their efficiency are poorly understood, and cosmological simulations are in general unable to drive these outflows directly without adopting various 'tricks' (see SD15 for a discussion). A frequently adopted assumption, loosely motivated by the expectations of momentum-conserving or energy-conserving winds (see SD15) is that the mass outflow rate scales with a power of the galaxy potential well depth, represented by the internal velocity dispersion or rotation velocity, times the star formation rate:

$$\dot{m}_{\rm out} = \epsilon_{\rm SN} \left(\frac{V_0}{V_{\rm c}}\right)^{\alpha_{\rm rh}} \dot{m}_*, \tag{2.3}$$

where $\dot{m}_*$ is the star formation rate, $V_{\rm c}$ is the circular velocity of the galaxy, normalised by the arbitrary constant $V_0 = 200\,{\rm km\,s^{-1}}$, and $\epsilon_{\rm SN}$ and $\alpha_{\rm rh}$ are tunable free parameters.

Variants of this recipe are almost universal in semi-analytic models, and several numerical cosmological hydrodynamic simulations effectively adopt it by administering 'kicks' to particles according to these or similar scaling relations (Oppenheimer & Davé 2006; Vogelsberger et al. 2014; Davé et al. 2016). Intriguingly, several simulations in which winds are driven by attempting to implement stellar feedback 'directly' also find that the wind mass loading factors $\eta \equiv \dot{m}_{\rm out}/\dot{m}_*$ follow this type of scaling. Christensen et al. (2016) find that their zoom-in simulations with the GASOLINE code and a 'blastwave' approach to driving winds yield $\alpha_{\rm rh} \sim 2$, and Muratov et al. (2015) find that at

high redshift η depends strongly on V_c in the FIRE simulations, where $\alpha_{\rm rh} \gtrsim 2$ for low-$V_{\rm c}$ halos and $\alpha_{\rm rh} \sim 1$ at higher $V_{\rm c}$.

2.2.5 Stellar Populations and Dust Attenuation

For each galaxy, we store a two-dimensional grid recording the mass of stars formed with a given age and metallicity. Unattenuated synthetic spectral energy distributions (SED) are created for each galaxy by convolving this grid with simple stellar population models (SSP). Here we use the SSP models of Bruzual & Charlot (2003) with the Padova1994 (Bertelli et al. 1994) isochrones and a Chabrier (2003b) initial mass function (IMF).

We include dust attenuation using a simple approach similar to the one presented in Somerville et al. (2012, hereafter S12). We assume that the face-on extinction optical depth in the V-band is given by

$$\tau_{V,0} = \tau_{\rm dust,0}\, Z_{\rm cold}\, m_{\rm cold}/(r_{\rm gas})^2, \tag{2.4}$$

where the optical depth normalisation of dust, $\tau_{\rm dust,0}$, is a free parameter, $Z_{\rm cold}$ is the metallicity of the cold gas, $m_{\rm cold}$ is the mass of the cold gas in the disc, and $r_{\rm gas}$ is the radius of the cold gas disc. We then calculate the attenuation in the V-band for a given inclination using a 'slab' model, which assumes that the radiation sources (stars) are embedded in a slab of dust:

$$A_V = -2.5 \log_{10} \left[\frac{1 - \exp[-\tau_{V,0}/\cos(i)]}{\tau_{V,0}/\cos(i)} \right]. \tag{2.5}$$

Here, i is the inclination, which is chosen randomly for each galaxy. We then translated this to the attenuation in the UV-band (1600Å), $A_{\rm UV}$, assuming a fixed attenuation curve (Calzetti et al. 2000). We experimented with using the two-component Charlot & Fall (2000)-type model (as implemented in S12), in which an age-dependent attenuation curve is effectively adopted by assuming that young stars are enshrouded in higher optical-depth 'birth clouds'. However, we found that this type of model predicted very large amounts of extinction in high redshift galaxies, which was incompatible with observations.

Several previous works have found that simple models for dust attenuation, in which the dust extinction optical depth scales with gas metallicity and column density with a fixed normalisation, overpredict the amount of attenuation in high redshift galaxies (S12; Wilkins et al. 2013; Reddy et al. 2015, 2018; Whitaker et al. 2017). One possible interpretation is that the dust-to-metal ratio is not constant, and depends on galaxy properties that effectively cause it to change with cosmic time. Indeed, recent observations have shown that the dust-to-metal ratio differs significantly across

different galaxy mass scales (Rémy-Ruyer et al. 2014), and models that attempt to self-consistently track the production and destruction of dust by various channels find that the dust-to-metal ratio may change over cosmic time (Popping et al. 2017b). We plan to develop more detailed and physical models of dust attenuation in high redshift galaxies in future work (Popping et al. in prep), but for the present work we adopt a very simple and ad hoc approach in which we allow the dust normalisation parameter $\tau_{\rm dust,0}$ to be a function of redshift (S12, see also: Inoue 2003; Guo & White 2009; Lo Faro et al. 2009; Santini et al. 2014; Wiseman et al. 2017). We tune $\tau_{\rm dust,0}$ by hand and find the values summarized in Table 2.2 provide a good match to available observational constraints on the bright-end of the UV LFs at $z = 4$–10. This is then implemented in our model with a functional form $\tau_{\rm dust,0} = \exp(az + b)$ with parameter values $a = -0.31$ and $b = -2.726$. A comparison of the values adopted in this work and in S12 can be found in Fig. 2.1. Note that the cosmological parameters and SAM ingredients and parameters adopted in S12 are a little bit different than the ones adopted in this work, which is why the values do not match up perfectly in the redshift range of overlap.

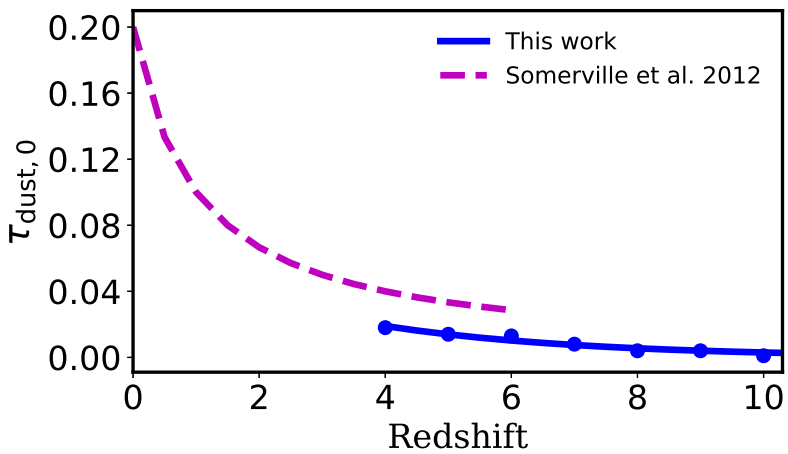

Figure 2.1 Comparison of the $\tau_{\rm dust,0}$ values adopted in this work and in S12. The blue markers show the values tuned by hand to match observations, and the solid line is the fitting function adopted in our model (see text).

We obtained the latest version of the published NIRCam filter response functions[1] with optical telescope element (OTE). For each galaxy, we calculate its apparent magnitude by convolving the synthesized SED with the NIRCam filters. We also include absorption due to the intervening IGM. The effective optical depth of the IGM along the line-of-sight, to a source at some redshift, at wavelength λ is calculated using the expression given in Madau et al. (1996).

[1] https://jwst-docs.stsci.edu/display/JTI/NIRCam+Filters

Table 2.2 Values for $\tau_{\rm dust,0}$ tuned by hand to match observations and fitted with the exponential function provided in the text.

	$\tau_{\rm dust,0}$ (10^{-3})	
z	Tuned	Fitted
4	18.00	18.99
5	14.00	13.93
6	13.00	10.22
7	8.00	7.50
8	4.00	5.50
9	4.00	4.04
10	1.00	2.96

2.2.6 Chemical Evolution

The production of metals is modelled using a simple approach that is commonly adopted in semi-analytic models (see e.g. Somerville & Primack 1999; Cole et al. 2000; De Lucia et al. 2004). In a given time-step where a parcel of new stars dm_* is created, a mass of metals $dM_Z = y\, dm_*$ is also formed, where y is the 'effective' chemical yield, or mean mass of metals produced per mass of stars. Here we assume that the chemical yield is constant. In principle, y could be obtained from stellar evolution models, but these model yields are uncertain by a factor of ~ 2, and the single-element instantaneous recycling approach to chemical evolution that we are using here is somewhat crude, so we instead treat the chemical yield as a free parameter while restricting it to an expected range. Once created, metals are assumed to be mixed instantaneously with the cold gas in the disc. We track the mean metallicity of the cold gas $Z_{\rm cold}$, and new star parcels created out of this gas are assumed to have the same stellar metallicity Z_* as the mean metallicity of the cold gas in that time-step. Supernova feedback ejects metals from the disc, along with cold gas. These metals are either mixed with the hot gas in the halo, or ejected from the halo into the 'diffuse' Intergalactic Medium (IGM), in the same proportion as the reheated cold gas(see S08). The ejected metals in the 'diffuse gas' reservoir are also reaccreted into the halo in the same manner as the gas.

Throughout this chapter, the yield y and all metallicities are given in solar units, which we take to be $Z_\odot = 0.02$. Although this formally represents the total metallicity, we note that as we track only the enrichment associated with Type II supernovae, our metallicity estimates probably correspond more closely with α-type elements. Note that because enriched gas may be ejected from the halo, and primordial gas is constantly being accreted by the halo, this approach is not equivalent to a standard 'closed box' model of chemical evolution.

2.2.7 Calibrating the free parameters

We calibrate our models to a standard set of $z \sim 0$ observables, and then leave all free parameters (except the dust normalisation, as noted above) fixed. Relative to the WMAP5 cosmology used in S08 and SPT15, the Planck cosmology adopted here results in significantly different predictions for the abundance of dark matter halos as a function of cosmic time (see Rodríguez-Puebla et al. 2016), and as a result the free parameters of the SAM need to be re-calibrated. We show the calibration quantities, along with some other diagnostic quantities that are not used in calibration, in Appendix B. The details of the calibration procedure are also presented in the Appendix. The model parameters used throughout this work are summarized in Table 2.3.

Table 2.3 A table for the model parameters that changed after re-calibration for the Planck cosmology. We also show the values used in SPT15, where cosmological parameters from WMAP5 were adopted, to illustrate by how much these parameters have changed. For a complete list of model parameters, see Table 1 in SPT15.

Parameter	Description	This work	SPT15
$\epsilon_{\rm SN}$	SN feedback efficiency	1.7	1.5
$\alpha_{\rm rh}$	SN feedback slope	2.8	2.2
$\tau_{*,0}$	SF timescale normalisation	1.0	1.0
y	Chemical yield (in solar units)	2.0	1.6
$\kappa_{\rm AGN}$	Radio mode AGN feedback	3.0×10^{-3}	3.8×10^{-3}

2.3 Rest-Frame UV Luminosity Functions

Perhaps the most basic statistical characterization of the galaxy population is the one-point distribution function of an observable quantity, such as luminosity (commonly referred to as the 'luminosity function' (LF)). The change of the LF in different redshift bins constrains the evolution of the galaxy population over cosmic time. In this section, we present the UV LFs predicted by our SAM at $z = 4 - 10$. With the large dynamic range of dark matter halo masses probed by our models, our predictions cover a wide UV luminosity range between $M_{\rm UV} \sim -6$ to -24. All binned luminosity functions presented in this work are available for download online[2]. Throughout this work we use a tophat filter of width of 400Å centred at 1600Å to calculate the rest-frame UV luminosity.

In Fig. 2.2, we show the distribution functions for the intrinsic rest-frame UV luminosities, without accounting for the effect of dust attenuation. Our results show that the choice of star formation recipe can significantly alter the number density of bright galaxies. Recall that in the GK-Big1 model, the star formation efficiency (molecular gas depletion timescale) is effectively fixed

[2]https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/

at a constant value. Keeping in mind that the H_2 depletion timescale in nearby spirals is 1–2 Gyr (Bigiel et al. 2008; Leroy et al. 2008), it is perhaps unsurprising that this becomes the limiting factor in forming stars at times when the age of the Universe is significantly less than this. In the GK-Big2 model, the 'steepening' of $N_{\rm SF} \rightarrow 2$ effectively leads to a higher star formation efficiency and shorter $t_{\rm dep,mol}$ in higher density gas. As galaxies are much more compact and gas rich at high redshift, this effectively leads to higher SF efficiencies at high redshift. The super-linear dependence of the 'classic' Kennicutt-Schmidt SF relation (KS) goes in the same direction, but to a lesser extent, as it assumes a slightly shallower slope $N_{\rm SF} = 1.5$. Interestingly, as already suggested in SPT15, we find that the formation of molecular gas is not a significant limiting factor for star formation even at these very high redshifts, in contrast to the suggestions of Krumholz & Dekel (2012). On the other hand, the faint end of the LF is insensitive to the choice of SF model (within the limited range of models that we have tested here). We will discuss the reasons for this in Section 2.5.

We show our results alongside with a compilation of UV LF constraints on the bright end presented in Finkelstein (2016), which consist of both ground- and space-based observations from McLure et al. (2009); Castellano et al. (2010); van der Burg et al. (2010); McLure et al. (2013); Oesch et al. (2013, 2014); Schenker et al. (2013); Tilvi et al. (2013); Bowler et al. (2014, 2015); Bouwens et al. (2015b, 2016b); Finkelstein et al. (2015a); Schmidt et al. (2014); McLeod et al. (2015, 2016). Note that the observational UVLFs are potentially impacted by attenuation by dust in these high-z galaxies, but dust attenuation is not included in the *intrinsic* model UV LF predictions shown in Fig. 2.2. We notice that the resultant LFs from the GK-Big1 model seem to fit observations fairly well while others overpredict the population of galaxies at $z \lesssim 6$ in the *absence* of dust. At higher redshifts $z \gtrsim 9$ the results from the KS and GK-Big2 models seem to in better agreement with observations.

In Fig. 2.3, we illustrate the effects of dust attenuation and photoionization squelching by comparing the outputs from our fiducial models with these effects switched on and off. The free parameter $\tau_{\rm dust,0}$ in our dust recipe is calibrated to match available constraints from *HST* observations by Bouwens et al. (2014a) and Finkelstein et al. (2015a). These observations along with constraints on the faint end from the Frontier Field by Livermore et al. (2017) are presented here for comparison. The dust model predicts that the effect of dust attenuation should be much stronger in bright, massive galaxies given their higher fraction of cold gas and the higher metallicity therein, whilst the effect on faint galaxies is minuscule. We fitted the UV LFs predicted by our fiducial model, with attenuation by dust, with the Schechter function using the least-squares method, where the

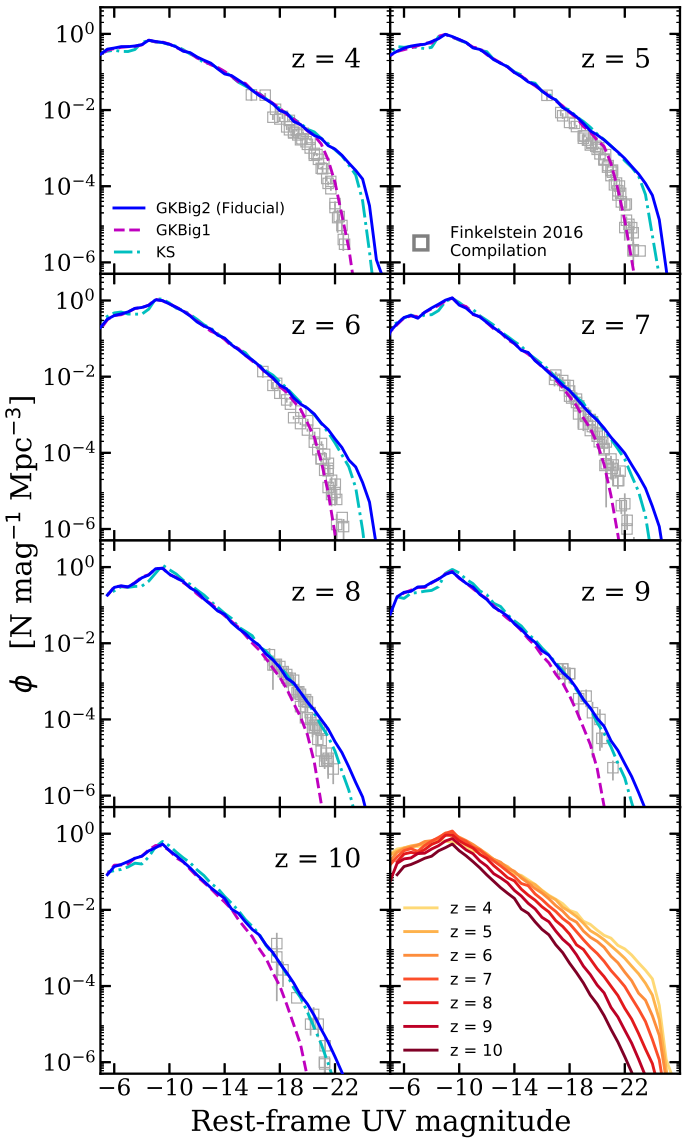

Figure 2.2 Predicted intrinsic UV LFs (without correction for dust attenuation) and their evolution with redshift. The blue solid line shows the results of the GKBig2 (fiducial) model, the purple dashed line shows the GKBig1 model, and the cyan dot-dashed line shows the KS model. We also include a compilation of observational constraints from Finkelstein (2016, squares) to guide the eye. The last panel summarizes the evolution predicted by the fiducial model.

Schechter function is given by

$$\phi(M) = 0.4\ln(10)\phi^{*}10^{-0.4(M-M^{*})(\alpha+1)}e^{-10^{-0.4(M-M^{*})}} \tag{2.6}$$

(Schechter 1976). The best-fit Schechter parameters are presented in Table 2.4.

Table 2.4 The best-fit Schechter parameters for the UV LFs inclusive of dust attenuation from our fiducial model between $z = 4 - 10$.

z	ϕ^{*} [10^{-3} Mpc^{-3}]	M^{*} [AB Mag]	α
4	3.151	-20.717	-1.525
5	2.075	-20.774	-1.602
6	1.352	-20.702	-1.672
7	0.818	-20.609	-1.715
8	0.306	-20.660	-1.825
9	0.133	-20.584	-1.879
10	0.053	-20.373	-1.967

In both cases, where dust attenuation is or is not included, our results show that the demographics of luminous galaxies seem to have evolved more rapidly than their faint counterparts. For instance, the 'knee' feature in the LFs is rather modest at high redshifts but this feature quickly develops and becomes distinct at $z \sim 7$. Moreover, we show that the position of the knee evolves as a function of redshift due to the rapidly evolving dust content. The extinction in the UV-band due to the presence of dust is estimated based on the physical properties of individual galaxies. The UV extinction as a function of $M_{\rm UV}$ is illustrated in Fig. 2.4. The two-dimensional histograms are colour-coded according to the conditional number density (Mpc^{-3}) of galaxies in each bin, which is normalised to the sum of the number density in its corresponding (vertical) rest-frame UV magnitude bin. The 50th, 16th, and 84th percentiles are marked in each panel to illustrate the statistical distribution. We have verified that the scatter in $A_{\rm UV}$ is dominated by the scatter in physical properties that are used to calculate $\tau_{V,0}$ (see Eqn. 2.4), whilst the randomly assigned inclination i is sub-dominant. Our results show that even with our simple approach to modelling dust, the scatter can be quite large.

Fig. 2.5 shows the $\beta_{\rm UV}$-$M_{\rm UV}$ relation, where the rest-frame UV luminosity presented here includes dust attenuation, and the photometric rest-frame UV spectral slope $\beta_{\rm UV}$ is calculated using the following expression

$$\begin{aligned}\beta_{\rm UV} &= \frac{\log(f_{\lambda,\rm FUV}/f_{\lambda,\rm NUV})}{\log(\lambda_{\rm FUV}/\lambda_{\rm NUV})} \\ &= -0.4\frac{(m_{\rm FUV} - m_{\rm NUV})}{\log(\lambda_{\rm FUV}/\lambda_{\rm NUV})} - 2\end{aligned} \tag{2.7}$$

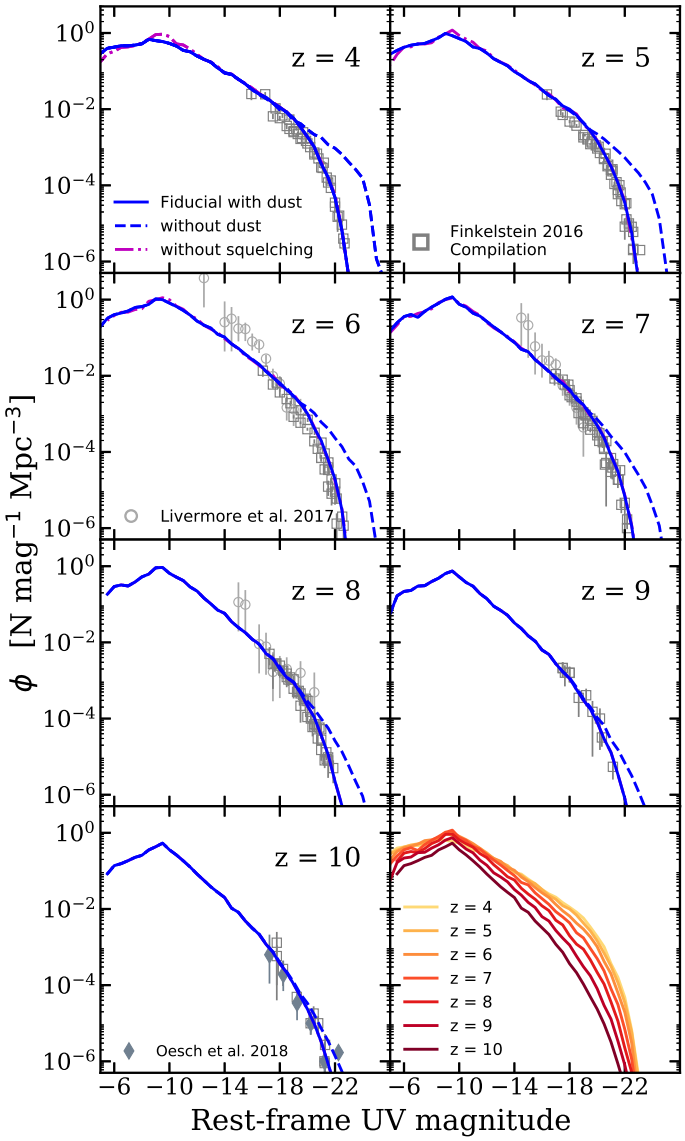

Figure 2.3 Redshift evolution of the dust-attenuated UV LFs between $z = 4 - 10$ predicted by our fiducial model (blue solid line). The blue dashed line shows the intrinsic UV LFs and the purple dot-dashed line shows the UV LFs without the effect of photoionization squelching, both from the fiducial model. We also include a compilation of observational constraints from Finkelstein (2016, squares) (same as Fig. 2.2) to guide the eye. Addition observational constraints from Livermore et al. (2017, circles) and Oesch et al. (2018, diamonds) are shown in $z = 6$, 7, 8, and 10. The last panel summarizes the evolution of the dust-attenuated UV LFs predicted by the fiducial model.

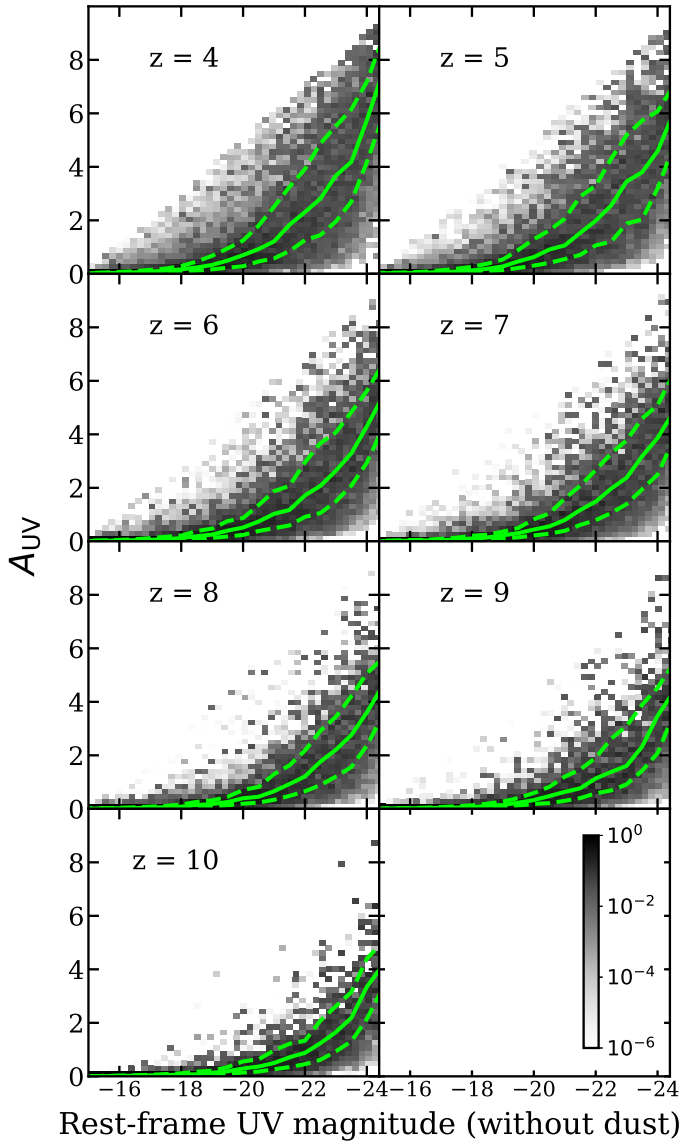

Figure 2.4 Conditional distributions of extinction in the UV-band versus intrinsic rest-frame UV magnitude in our fiducial model between $z = 4$–10. The green solid and dashed lines show the 50th, 16th, and 84th percentiles. The two-dimensional histograms are colour-coded to show the conditional number density of galaxies in each bin, which is normalised to the sum of the number density (Mpc^{-3}) in its corresponding (vertical) rest-frame UV magnitude bin.

from Onodera et al. (2016), where $\lambda_{\rm FUV}$ and $\lambda_{\rm NUV}$ are the central wavelengths of the far- and near-UV bands from the *Galaxy Evolution Explorer* (*GALEX*) survey, where we adopted $\lambda_{\rm FUV} \simeq$ 1530 Å and $\lambda_{\rm NUV} \simeq 2300$ Å for our calculations. Our results are compared to the range of $\beta_{\rm UV}$ spanned by typical local starburst galaxies and the median values of $\beta_{\rm UV}$ from a compilation of observations, including GOODS-S DEEP, GOODS-S WIDE, HUDF09, and WFC3 Early Release Science (Koekemoer et al. 2011; Grogin et al. 2011; Bouwens et al. 2010; Oesch et al. 2010; Windhorst et al. 2011). We also compare our results to measurements reported by Dunlop et al. (2013) at $z = 7$–9, Bouwens et al. (2014b) at $z = 4$–8, Rogers et al. (2014) at $z = 5$, and Wilkins et al. (2016a) at $z = 10$. The predicted $\beta_{\rm UV}$-$M_{\rm UV}$ relation is surprisingly good agreement with the observations.

In Fig. 2.6, we zoom into the bright end of the UV LFs and explore the sensitivity of the bright end behaviour to the SF timescale $\tau_{*,0}$, where larger $\tau_{*,0}$ means less efficient star formation, and vice versa. This parameter also effectively multiplies the gas consumption timescale, and therefore larger $\tau_{*,0}$ would result in higher gas fractions (again, see White et al. 2015). Here we increase or decrease $\tau_{*,0}$ from our fiducial value by a factor of 2 and find that this results in galaxy populations with $M_{\rm UV} \gtrsim -18$ mildly deviating from our fiducial model, with the strength of the deviation seeming to scale with luminosity. Joint constraints on the gas and dust content and SFR in high redshift galaxies from ALMA and *JWST* will be extremely valuable for breaking the degeneracy between dust and star formation efficiency. With regard to the bright end, we also note that although AGN feedback is responsible for shaping the bright end of the galaxy luminosity function in our models at $z \lesssim 2$, we have checked that switching the AGN feedback on and off has no noticeable effect on any of the results presented here.

2.3.1 The faint galaxy populations

Fig. 2.7 shows a zoomed-in figure that focuses on the faint end of the UV LFs. Note that the effect of dust is predicted to be negligible in this regime due to the generally low metallicity in these galaxies. The turnover in the LF near $M_{\rm UV} \sim -9$ is not due to resolution, but corresponds to the sharp cutoff in the atomic cooling function at 10^4 K (which corresponds to $V_{\rm vir} \simeq 17$ km s^{-1}). The slope of the UV LFs remains fairly steep down to this limit. In addition to the estimates of the observed LF from the lensed Frontier Fields, we show constraints on the LF at $z = 4 - 7$ from multiple studies. Weisz et al. (2014) estimate the evolution of the faint end of UV LFs by measuring the star formation histories of 37 Local Group galaxies out to $z \sim 5$ using 'fossil evidence' from resolved stellar populations. Interestingly, their estimates are a factor of ~ 10 higher than our predictions. The 'fossil' method provides a very interesting complementary approach for estimating the faint end of the high redshift luminosity function, but currently it involves some rather uncertain corrections

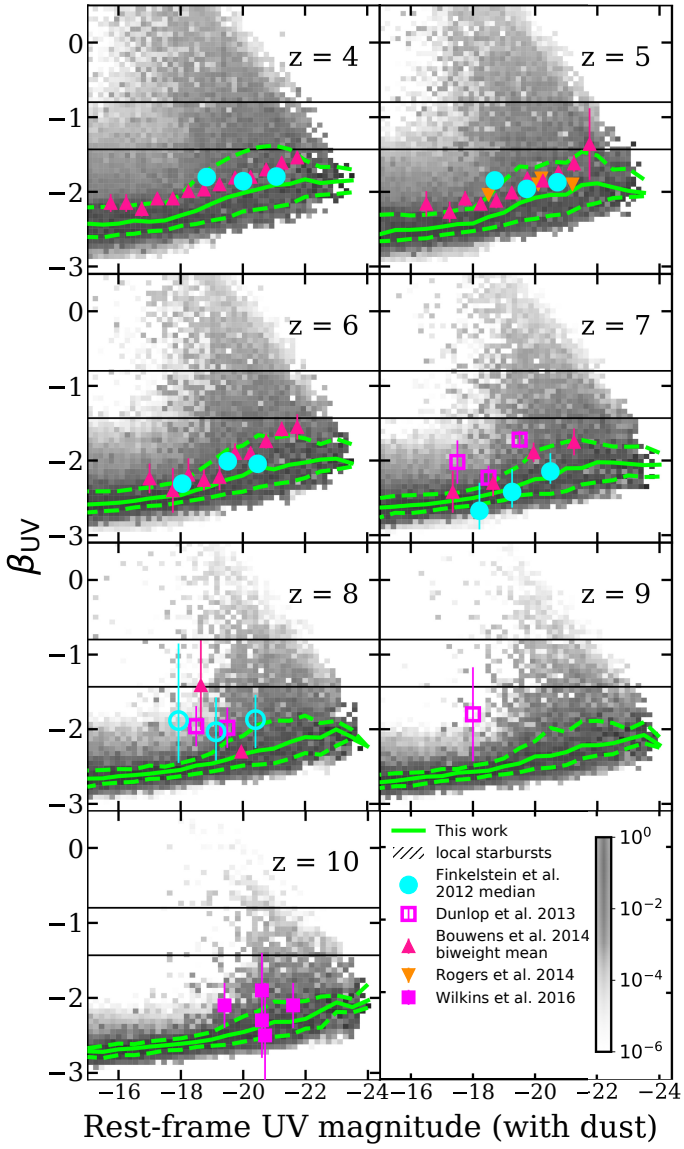

Figure 2.5 Conditional distributions of the rest-frame UV spectral slope $\beta_{\rm UV}$ versus rest-frame UV magnitude (with dust attenuation) in our fiducial model between $z = 4$–10. The green solid and dashed lines show the 50th, 16th, and 84th percentiles. The two-dimensional histograms are colour-coded to show the conditional number density of galaxies in each bin, which is normalised to the sum of the number density (Mpc^{-3}) in its corresponding (vertical) rest-frame UV magnitude bin. Data points are $\beta_{\rm UV}$ measurements reported by Finkelstein et al. (2012b, cyan circle), Dunlop et al. (2013, open pink square), Bouwens et al. (2014b, red triangle), Rogers et al. (2014, yellow square), and Wilkins et al. (2016a, pink square). The hatched gray band marks where typical local starburst galaxies lie (Finkelstein et al. 2012b, see text for details).

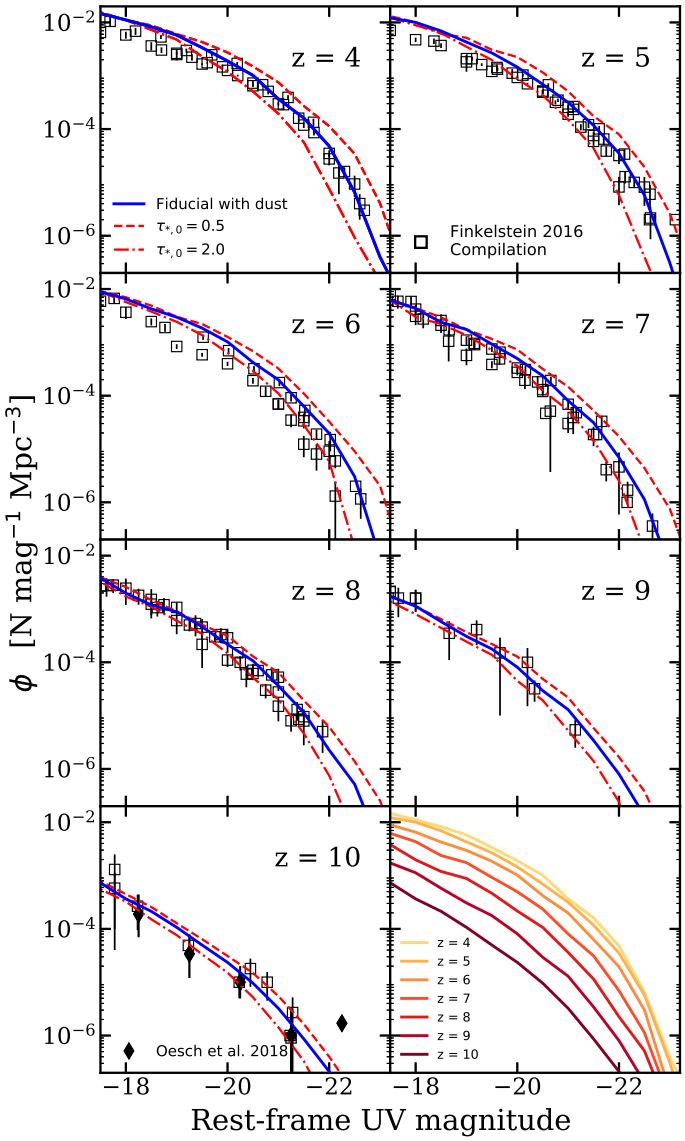

Figure 2.6 Redshift evolution of the bright end of the UV LFs between $z = 4 - 10$ predicted by our fiducial model with dust attenuation included (blue solid line). Black square markers represent a compilation of observational estimates from space- and ground-based surveys presented in Finkelstein (2016). Black diamond markers show the additional constraints at $z = 10$ from Oesch et al. (2018). Red lines represent the cases where we increase or decrease $\tau_{*,0}$ by a factor of 2; dashed and dot-dashed lines are $\tau_{*,0} = 0.5$ and $\tau_{*,0} = 2.0$, respectively. The last panel summarizes the evolution of the bright end of the dust-attenuated UV LFs predicted by the fiducial model.

(D. Weisz, private communication). The largest uncertainties likely arise from two factors. The first is the volumetric correction, which attempts to correct for the fact that star formation histories were not available for all known Local Group galaxies, and for Local Group galaxies that are not currently detected. The second issue is that deep resolved colour magnitude diagrams (that reach the oldest main sequence turnoff) were not at the time available for many Local Group galaxies. This can lead to a bias that yields increased SF estimates at early times (D. Weisz, private communication). Establishing better links between high-redshift observations and those derived from Local Universe galaxies with resolved stellar population studies is an exciting ongoing area of research, which *JWST* will also help to advance.

At $z \sim 6$–8, we show additional studies based on the Frontier Fields. Here we show both the published data from Livermore et al. (2017) and the unpublished, Eddington bias-corrected number densities for comparison (R. Livermore, private communication). Bouwens et al. (2017) provide constraints on $z \sim 6$ galaxies by reanalyzing the Frontier Field observations with a more comprehensive treatment for the magnification and systematic uncertainties. Atek et al. (2015) estimated constraints for LFs using a combined analysis of three lensed fields with their associated parallel fields.

In addition to the fiducial model parameters, we explored the sensitivity of the faint-end slope to the efficiency of stellar driven winds by comparing four additional cases of SN feedback slope $\alpha_{\rm rh} = 2.0$, 2.4, 3.2, and 3.6, where larger values of $\alpha_{\rm rh}$ imply a steeper dependence of the mass loading factor η on galaxy circular velocity, so more gas is ejected from low-mass galaxies. Although these values of $\alpha_{\rm rh}$ are not consistent with observations at $z \sim 0$, we present these results as an attempt to quantify the effects of strong stellar feedback on the low-mass galaxy populations and to illustrate how our predictions would change if the effective scalings vary with cosmic time. The results show that the faint end is indeed sensitive to the velocity scaling of the mass loading factor. Higher $\alpha_{\rm rh}$ leads to stronger suppression of star formation in low-mass halos, and hence to a flattening of the luminosity function. The adopted value of $\alpha_{\rm rh}$ also shifts the luminosity where the LF 'turns over' from ~ -8 (for the 'strongest') feedback to ~ -11 (for the weakest), because $\alpha_{\rm rh}$ also affects the slope of the relationship between luminosity (or stellar mass) and halo mass (so the luminosity of galaxies forming in halos with mass corresponding to the atomic cooling limit is shifted up or down). Note that varying the parameter $\epsilon_{\rm SN}$ simply shifts the normalisation of the whole luminosity function below the knee up or down, as shown in White et al. (2015).

Fig. 2.8 shows the evolution of the cumulative number density of galaxies above some rest-frame luminosity threshold $M_{\rm UV,lim}$. First we focus on the evolution of the bright end of the UV LFs, and present the results calculated with $M_{\rm UV,lim} = -19$. We show the bracketing cases from

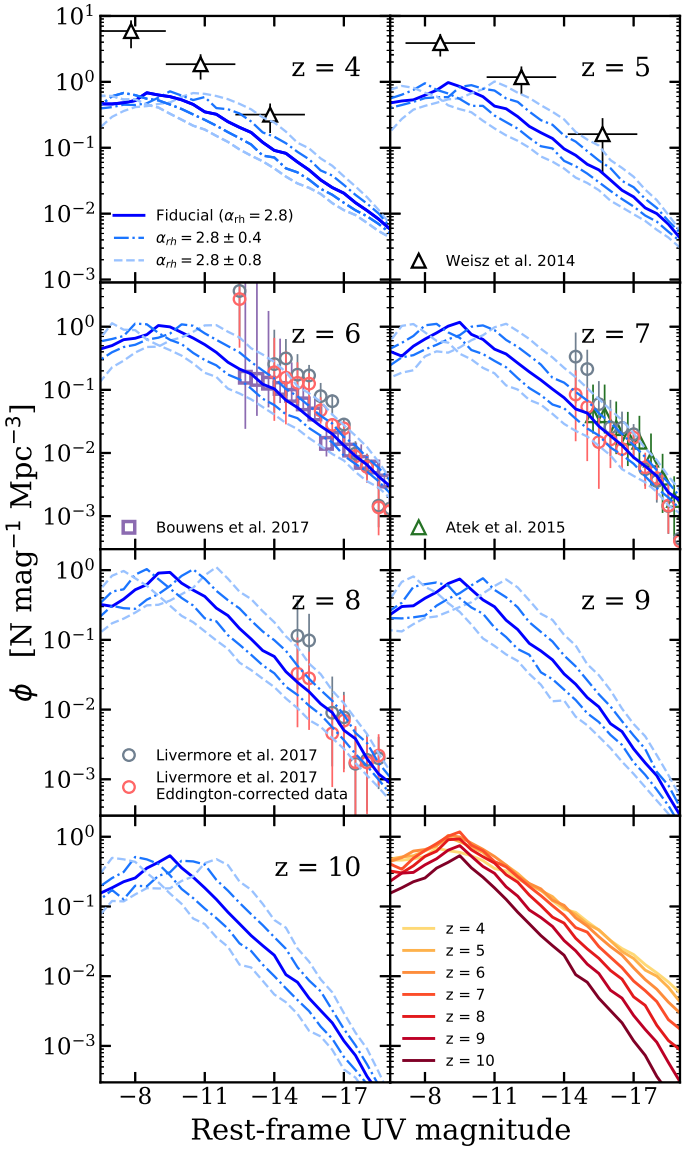

Figure 2.7 Redshift evolution of the faint end of the UV LFs between $z = 4 - 10$ predicted by our fiducial model with dust attenuation included (blue solid lines). Markers represent the observational estimates from Livermore et al. (2017) as originally published (grey circles) and with Eddington correction (red circles), as well as an independent analysis by Bouwens et al. (2017, purple squares) for $z \sim 6$. We also include estimates from Weisz et al. (2014, black triangle) and Atek et al. (2015, green triangle) for $z \sim 7$ (see text for details). We show four additional scenarios where we vary the parameter controlling the mass-loading of stellar driven winds, $\alpha_{\rm rh} = 2.8 \pm 0.4$ and ± 0.8. Blue dot-dashed lines show the cases where we let $\alpha_{\rm rh} = 2.4$ (above) and 3.2 (below), and light blue dashed lines show the cases where we let $\alpha_{\rm rh} = 2.0$ (above) and 3.6 (below). The last panel summarizes the evolution of the faint end of the dust-attenuated UV LFs predicted by the fiducial model.

our fiducial and GK-Big1 models, which provide the most and least optimistic scenario regarding forming luminous galaxies at high redshift. We compare our results to a compilation of studies from observations, empirical extrapolations, and numerical hydrodynamic simulations (Mason et al. 2015; Atek et al. 2015; Bouwens et al. 2015b; Ishigaki et al. 2018; Wilkins et al. 2017). This comparison shows that results from other studies are more or less bracketed by our two star formation scenarios.

Additionally, we probe the faint end of the LFs by showing the cumulative number density for galaxies with $M_{\rm UV} < -15.0$. Aside from our fiducial model, we include two bracketing cases where we let $\alpha_{\rm rh} = 2.0$ and 3.6, which correspond to weaker and stronger stellar feedback efficiency, respectively. This kind of test will place constraints on the efficiency of stellar winds and whether there are additional dependencies in the wind mass loading scalings beyond the simple assumptions adopted here. The comparison shows that the prediction from our fiducial model is similar to other studies. However, some studies predict fewer galaxies than we predicted at lower redshifts due to a shallower faint-end slope of the UV LFs, which in our model framework would require stronger stellar feedback.

In Fig. 2.9, we show the cosmic star formation rate (CSFR) for our fiducial and GK-Big1 models, with star formation rate integrated down to $M_{\rm UV} = -17 \pm 1.0$. Our result is compared to Behroozi et al. (2013), Finkelstein et al. (2015a), Bouwens et al. (2015b), Bouwens et al. (2016b), McLeod et al. (2016), and Ishigaki et al. (2018). The CSFR predicted by the KS model is very similar to the ones from our fiducial model, as the UV LF results hinted, and we therefore omitted the KS model from this figure.

Overall, it is intriguing that our models, which were previously calibrated and tested only at much lower redshifts, agree so well with the existing observations all the way out to $z \sim 10$. Our models predict that the number density of galaxies evolves quite rapidly between $z = 4 - 10$, and that the bright end evolves more rapidly than the faint end. Our models predict that the slope of the UV LFs will remain fairly steep below the current detection limit until $M_{\rm UV} \sim -9$. This slope is not sensitive to the choice of star formation model but is very sensitive to the scaling of the mass loading factor of stellar-driven winds with galaxy circular velocity, as is the luminosity where the LF turns over at the faint end. The bright end is sensitive to the star formation efficiency or gas depletion timescale and its dependence on gas surface density. However, the effects of dust and the SF efficiency on the bright end of the LF are degenerate, so independent probes of these quantities are needed.

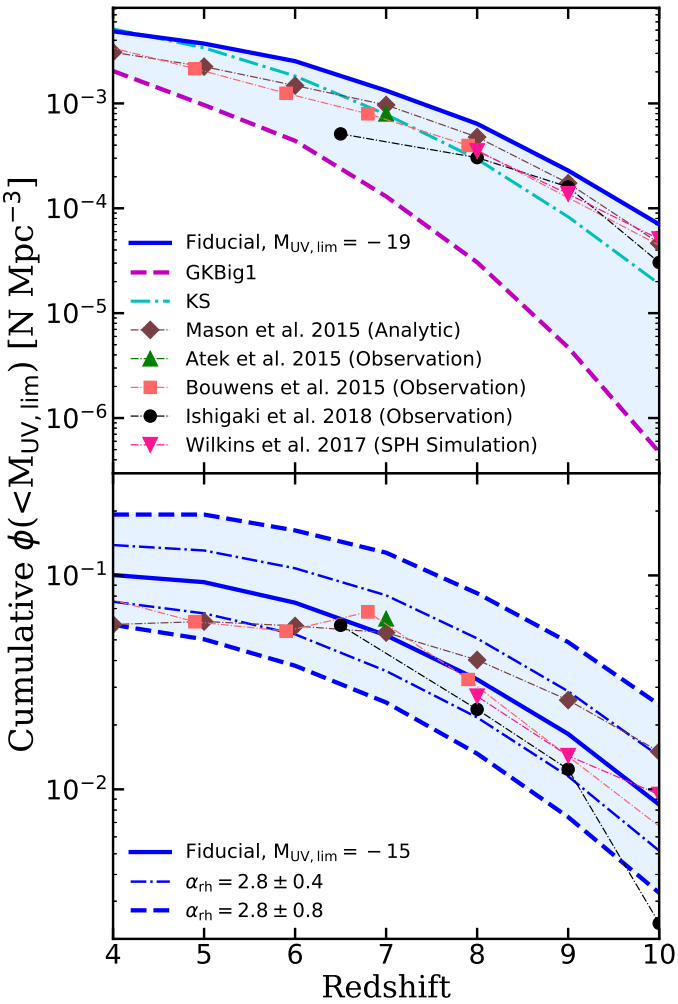

Figure 2.8 The cumulative number density for objects brighter than the specified rest-frame UV luminosities, $M_{\rm UV,lim} < -19$. The *upper panel* focuses on the evolution on the bright end. The blue and purple lines show the fiducial and GK-Big1 models, or our most and least optimistic scenarios, respectively. The dashed cyan line shows the result from the KS model. The *lower panel* focuses on faint objects with $M_{\rm UV,lim} < -15$. The blue solid line shows the fiducial model. The dot-dashed and dashed lines show the cases where we let $\alpha_{\rm rh} = 2.4$ and 2.0 (above), and $\alpha_{\rm rh} = 3.2$ and 3.6 (below), respectively. In both panels, the results are compared to a compilation of UV LFs from an SPH simulation, observations, and empirical extrapolations (Mason et al. 2015; Atek et al. 2015; Bouwens et al. 2015b; Wilkins et al. 2017; Ishigaki et al. 2018).

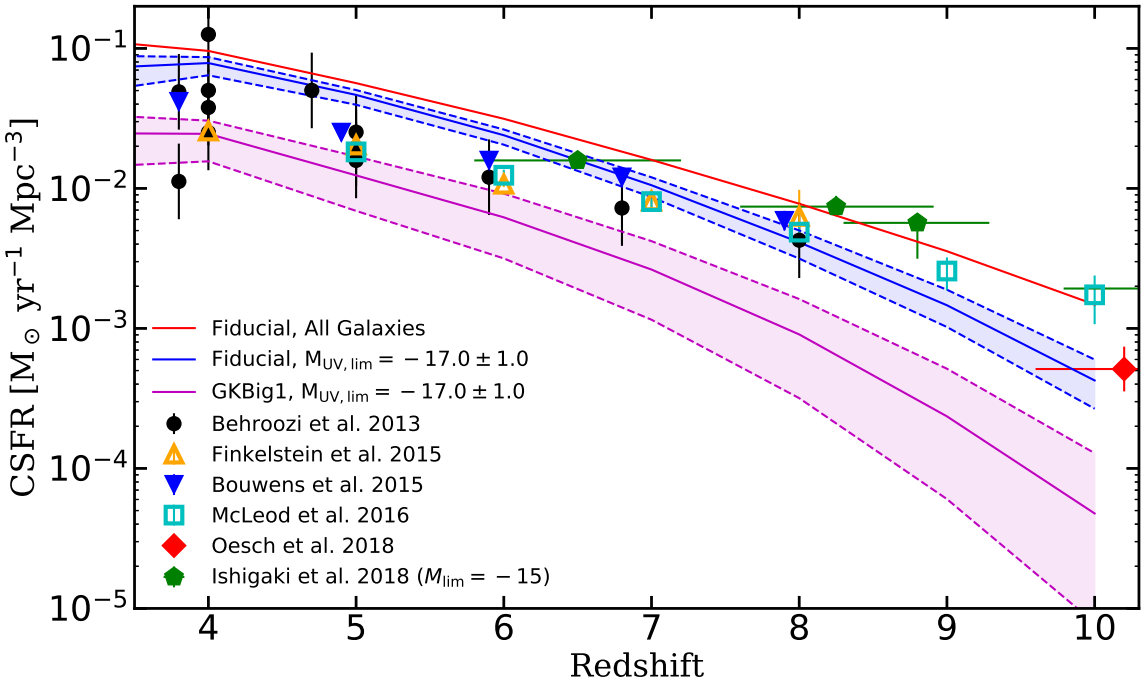

Figure 2.9 Cosmic star formation rate density evolution with redshift integrated down to $M_{\rm UV,lim} = -17.0 \pm 1.0$. The blue and purple lines show results from our fiducial and GKBig1 models, respectively. Our results are compared to observations from other studies by Behroozi et al. (2013, black circles), Finkelstein et al. (2015a, orange open triangles), Bouwens et al. (2015b, blue inverted triangles), McLeod et al. (2016, cyan open squares), Oesch et al. (2018, red diamonds), Ishigaki et al. (2018, green pentagons). The cut-off magnitude for these observations is $M_{\rm UV} = -17$, unless specified otherwise.

2.3.2 Comparison with other models

In addition to comparing our results with observational constraints, we also compare our predictions with a collection of theoretical studies, including empirical models, semi-analytic models, and cosmological hydrodynamic simulations. Note that studies that are fully or partially numerical are subject to the inevitable tension between simulated volume and spatial and time resolution. For example, simulations with high resolution are only feasible to run over small volumes and rare, massive objects are not well sampled. On the other hand, simulations with coarser resolution and larger volumes do well at capturing formation of large-scale structure. However, reliable predictions for small objects that fall near or below the resolution limit are not possible. For this reason, the galaxy mass and luminosity range covered varies among these studies. And note that among all models compared here, our model has the widest dynamic range and is capable of carrying the simulation to $z = 0$ for comparison with local observations.

Overall, our results agree extremely well with other studies in regimes where observational constraints are available. On the other hand, our predictions for faint galaxy populations, for which observational constraints are very limited due to the lack of direct detections, are in close proximity with a number of other studies, so long as the simulations have adequate resolution to resolve these objects. The faint-end slopes predicted by our fiducial model tend toward the shallower end, especially at higher redshifts, among the range collectively predicted by these models.

Both empirical models included in this comparison employ some empirical relations between halo mass and star formation efficiency that are calibrated using halo abundance matching at redshifts where observational constraints are available. These relations are then extrapolated to make predictions at higher redshifts. The Mason et al. model, which is based on models introduced by Trenti et al. (2010, 2015); Tacchella et al. (2013), uses a redshift independent star formation efficiency that depends on the halo mass and assembly time, calibrated at $z \sim 5$. The Sun & Furlanetto study explores the outcome of various star formation efficiency models, including one that assumes a power-law extrapolation below a cutoff halo mass limit $M_{\rm h} = 2 \times 10^{10}$ $\rm M_\odot$ and a best-fit redshift independent model. Note that their models are calibrated to bright galaxies at $z = 6$–8, and results at $z = 9$ and 10 are predictions that are yet to be published.

On the other hand, the details for the numerical hydrodynamic simulations included in this comparison are summarized in the following. BlueTides is a large-volume cosmological hydrodynamic simulation that focuses on the high-redshift ($z \gtrsim 8$) universe, with a box that is 400 Mpc h^{-1} on a side, resolving galaxies with $M_* \gtrsim 10^8$ $\rm M_\odot$ toward the end of their simulation (Wilkins et al. 2017). Vulcan is a high-resolution (dark matter particle $\sim 10^5$ $\rm M_\odot$), relatively small volume (25 comoving

Mpc on a side) simulation that aims to quantify the contribution of faint galaxies to cosmic reionization (Anderson et al. 2017). The Cosmic Dawn (CoDa) simulation is a large-scale hydrodynamic simulation coupled with radiative transfer modelling, resolving galaxies down to $\sim 10^8$ $M_\odot$ in box that is ~ 100 Mpc on a side (Ocvirk et al. 2016). Cosmic Reionization On Computers (CROC) is a cosmological simulation based on the adaptive refinement tree (ART) method (Gnedin 2014; Gnedin & Kaurov 2014). The CROC results compared here are from their fiducial simulation box with 20 Mpc h^{-1} on a side and 512^3 initial grid cells, with spacial resolution of 100 pc (see Gnedin 2016). The DRAGONS project (Poole et al. 2016) consists of a SAM MERAXES that is built on top of the *Tiamat* suite of n-body simulations. We also include results from a high-resolution hydrodynamic simulation with embedded self-consistent radiative transfer that is similar to the one described in Finlator et al. (2012, 2015, 2016, 2017). The three runs included here have box sizes of 6.0, 7.5, and 10.0 Mpc h^{-1} on a side, with 2×256^3, 2×320^3, and 2×512^3 particles; the UV background is discretized into 32, 40, and 64 voxels.

In Fig. 2.10, we compare the bright end of our UV LFs to the predictions of Mason et al., Sun & Furlanetto, DRAGONS, CROC, CoDa, VULCAN, and BLUETIDES. In Fig. 2.11, we compare our predictions for the faint-end behaviour of the UV LFs to Mason et al., Sun & Furlanetto, DRAGONS, CROC, and Finlator et al.. The Mason et al. model assumed a $M_{\rm UV} = -12$ cutoff for the atomic cooling limit in their work and the Sun & Furlanetto model considered a limiting magnitude of $M_{\rm UV} = -13$ in their CSFR calculation. For this comparison, we plot the results from these analytic studies extrapolated down to fainter magnitudes to illustrate the faint-end slopes predicted by these models and how they stack up with other numerical results. On the other hand, numerical simulations provide predictions with physical processes traced self-consistently as far down as resolution permits, although they are still limited by the assumptions inherent in their sub-resolution recipes for star formation and stellar feedback. Moreover, we note that the very faint populations predicted by these numerical studies might be somewhat resolution dependent, rather than physical. For instance, some subtle differences can be spotted between the UV LFs based on the coarser *Tiamat* (at $z = 6$–10) and the finer *Tiny Tiamat* (at $z = 6$, 8, and 10) N-body simulations from the DRAGONS simulation suite in our comparison (Liu et al. 2016). Similar behaviour can be seen among the three different runs in the Finlator et al. simulations. However, heating from a photoionizing background is tracked self-consistently with on-the-fly radiative transfer in this model. The *Renaissance Simulations* (O'Shea et al. 2015), SPHYNX (Cabezón et al. 2017), and FIRE (Ma et al. 2018b) are theoretical studies that address similar issues but are not included in our comparison here.

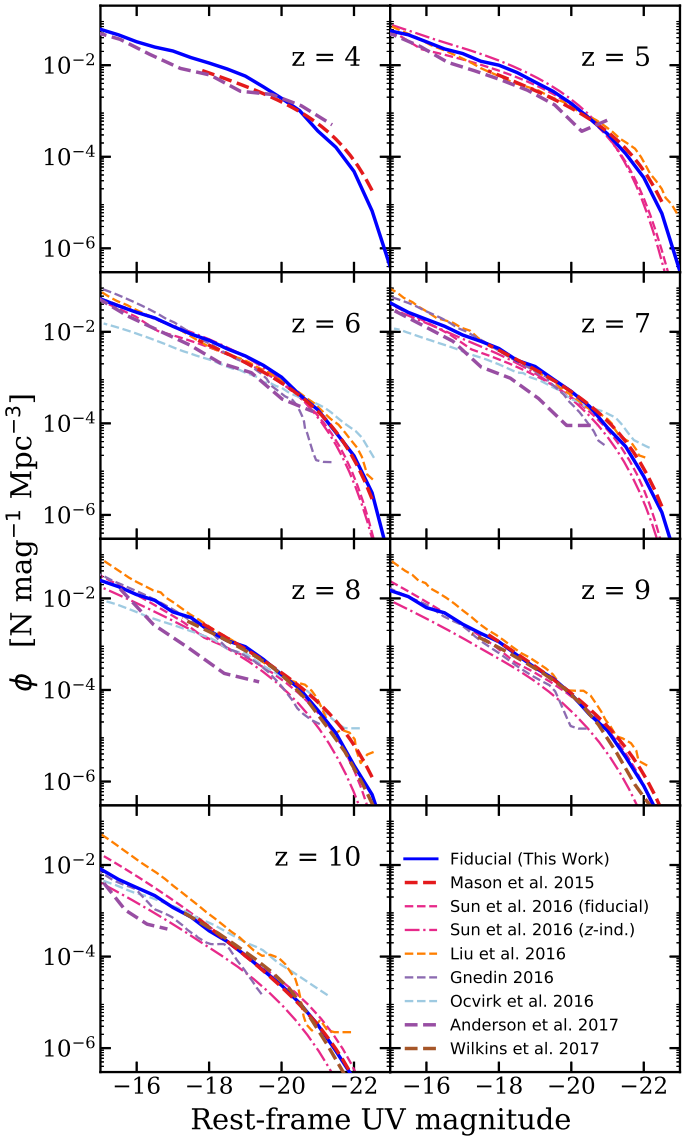

Figure 2.10 Redshift evolution of the bright end of the UV LFs between $z = 4$–10 predicted by our fiducial model with dust attenuation (blue solid lines). Our results are compared to several other theoretical studies, including empirical studies from Mason et al. (2015, green) at $z = 4$–10 and Sun & Furlanetto (2016, red) at $z = 5$–10 from their fiducial model (dashed) and redshift-independent model (dot-dashed), cosmological hydrodynamic simulations BlueTides (Wilkins et al. 2017, brown) at $z = 8$–10 and Vulcan (Anderson et al. 2017, purple) at $z = 4$–10, and the DRAGONS SAM simulation suite (Liu et al. 2016, orange) at $z = 5$–10. See text for details.

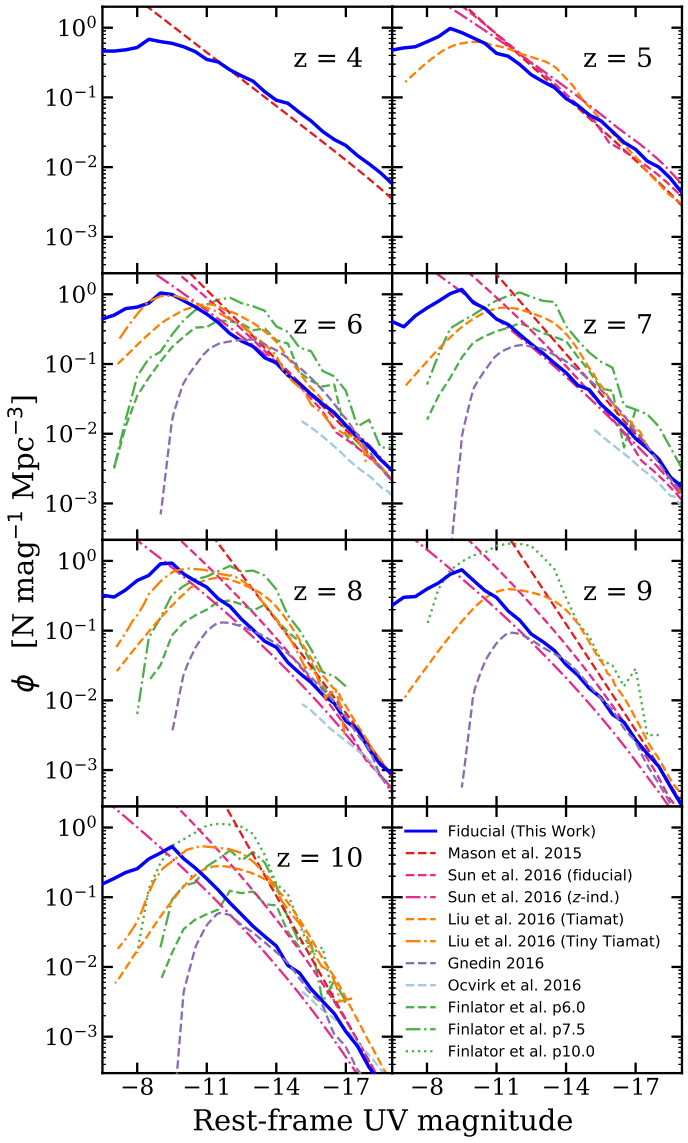

Figure 2.11 Redshift evolution of the faint end of the UV LFs between $z = 4$–10 predicted by our fiducial model with dust attenuation (blue solid line). Our results are compared to several other studies, including empirical models from Sun & Furlanetto (2016, red) at $z = 5$–10 fiducial model (dashed) and z-independent model (dot-dashed), DRAGONS based on their *Tiamat* (dashed) and *Tiny Tiamat* (dot-dashed) simulations at $z = 5$–10 (Liu et al. 2016, orange). We also include results from a hydrodynamic simulation with embedded self-consistent radiative transfer that is similar to the one presented in Finlator et al. (2012, green) at $z = 6$–10 in three different box sizes and resolutions (dashed, dot-dashed, and dotted in increasing resolution). See text for more details.

2.4 Predicted Apparent Magnitude Functions with NIRCam Filters

In this section, we provide the distribution functions of apparent magnitude (AMF) for high-redshift galaxy populations, where the latest published *JWST* NIRCam filter response functions are used to compute the apparent magnitude for the galaxies predicted by our model. We also provide estimates of the number of detected objects for several example survey configurations.

In Fig. 2.12, we showcase the evolution of the AMFs over redshift as seen in the eight NIRCam broadband filters: F070W, F090W, F115W, F150W, F200W, F277W, F356W, and F444W. The central wavelength and the detection limit of these filters are summarized in Table 2.5. Detection limits are 10-σ point source limiting AB magnitudes from the *JWST* Exposure Time Calculator[3] for an exposure time of 10^4 second, measured in 0.04' diameter apertures for the short-wavelength camera, and 0.08' diameter apertures for the long-wavelength camera, assuming a low background level. An approximate detection limit $m_{\rm lim} \sim 29$ is marked in the figure with a black dashed line. Objects to the left of the line are too faint to be detected at this depth. We also marked where one object is expected per NIRCam field of view ($2 \times 2.2 \times 2.2$ arcmin^2) with a horizontal dashed line in each panel. Objects with counts below this line are too rare to be found on average in a single pointing of NIRCam. The seven panels together demonstrate how the high-redshift galaxy populations quickly drop out from the filters with shorter wavelengths due to absorption by the intervening IGM. Our results show that the F200W filter, our choice of nominal filter, is best suited for detecting objects across all redshifts of interest, and might even be able to pick up bright objects beyond $z \sim 10$ with an extended exposure time. The Schechter function fitting parameters for the F200W filter are presented in Table 2.7. Redundant detections from multiple filters also improve the accuracy of estimates of dust attenuation, which is essential to properly uncover the underlying intrinsic UV luminosity. Galaxies at high redshifts rapidly drop out of the shorter wavelength filters F070W, F090W, and F115W for $z < 10$, and F150W at $z \sim 12$, whilst F200W, F277W, F356W, and F444W remain advantageous for possible detections up to $z \sim 15$ given sufficient exposure time. Although MIRI is designed to detect longer wavelength radiation, it will not be able to detect such high redshift galaxies due to its low sensitivity.

In Fig. 2.13, we further investigate the population of galaxies that are expected to be found by *JWST* in different surveys. We show the same AMFs from the F200W filter and provide the detection limits from a number of simulated surveys. The detection limits of past *HST* surveys and upcoming *JWST* surveys are summarized in Table 2.6. First, we show the detection limit of the *HST* CANDELS-Wide survey for reference. Then, from bright to faint, we show the detection limits

[3] https://jwst.etc.stsci.edu/, v1.2.2

estimated for a *JWST* investment comparable to CANDELS-W and HUDF, and the very optimistic case where we gain a factor of 10× magnification from gravitational lensing from a massive galaxy cluster. We indicated where ~ 10 objects are expected to be found in survey areas similar to that of the HUDF field ($2.4 \times 2.4\text{arcmin}^2$) and CANDELS-W ($\sim 100$ arcmin^2), which are both estimated assuming $dz = 1$ slices centred at the given redshift. We also show the cases where we deviate from the fiducial value and let $\alpha_{\rm rh} = 2.8 \pm 0.4$ and ± 0.8, where stronger feedback gives shallower faint-end slopes and vice versa. This comparison shows that *JWST* will be able to reliably constraint the faint-end slope up to $z \sim 8$, or even $z \sim 10$ in lensed fields. This will provide important constraints on stellar feedback physics.

In fig.2.14, we show the number of objects per Mpc^3 above some detection limits, $m_{\rm lim}$, expected to be found using the NIRCam F200W filter. We consider two cases where we let $m_{\rm lim,wide} = 29$ and $m_{\rm lim,deep} = 31.5$, representing the expected detection limit for *JWST* wide-field surveys (e.g. CANDELS-Wide) and deep-field surveys (e.g. HUDF). In the upper panel, we illustrate the difference among the three SF models considered in this work, and in the lower panel we show the cases where we let $\alpha_{\rm rh} = 2.0$ and $\alpha_{\rm rh} = 3.6$. Here, a more direct comparison with observations can be made, without the need for applying K-corrections to the observations to get from observed to rest-frame magnitudes. One can see that, similarly, observations of the bright population will constrain the star formation efficiency, while observations of the faint end will constrain stellar driven winds.

Table 2.5 Pivot wavelengths and bandwidths of NIRCam filters. Detection limits assuming a 10^4 seconds exposure. [a] *JWST* User Documentation[4]

NIRCam Filters	λ^a [μm]	Bandwidth^a [μm]	Detection Limit [AB Mag]
F070W	0.704	0.132	28.16
F090W	0.902	0.194	28.56
F115W	1.154	0.225	28.85
F150W	1.501	0.318	29.04
F200W	1.989	0.457	29.07
F277W	2.762	0.683	28.93
F356W	3.568	0.781	28.93
F444W	4.408	1.029	28.33

2.5 Discussion

The formation and evolution of galaxies are governed by a complex network of intertwined physical processes that operate over many physical scales. Although many details remain to be worked out, the community seems to have reached a broad consensus regarding the main physical processes

[4] https://jwst-docs.stsci.edu/display/JTI/NIRCam+Filters

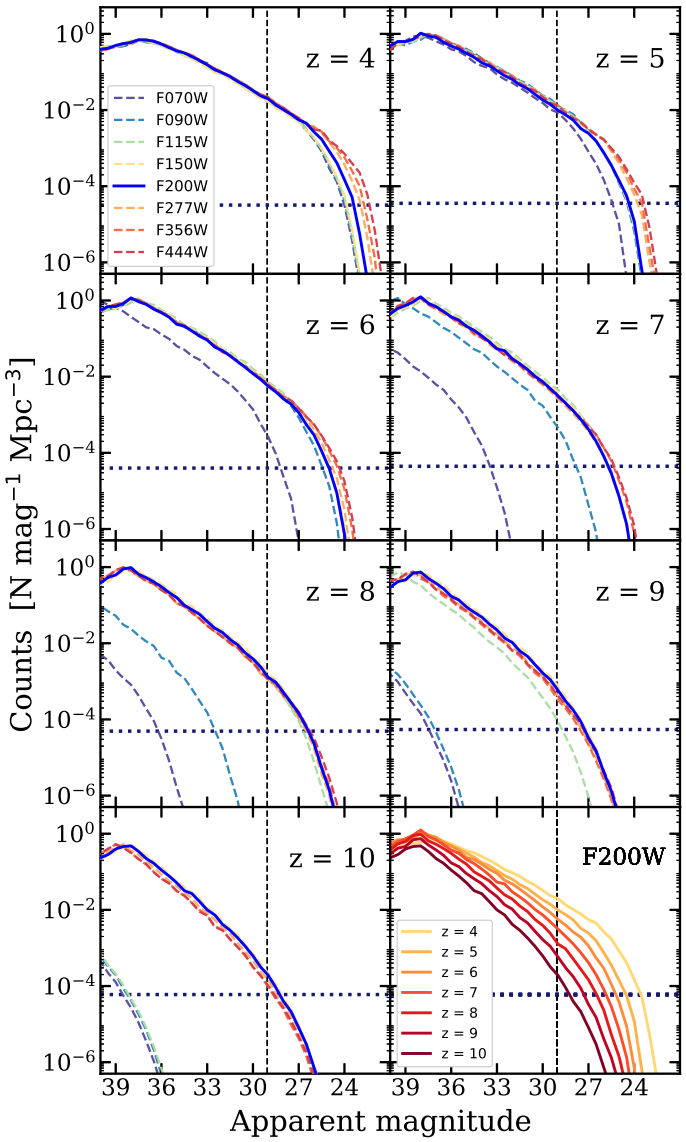

Figure 2.12 Redshift evolution of the AMFs as seen in every NIRCam filter between $z = 4 - 10$. See Table 2.5 for the specifications for the filters compared here. The vertical black dashed lines represents the detection limit of NIRCam assuming a 10^4 second exposure. The horizontal dashed line shows where one object is expected per NIRCam field of view ($2 \times 2.2 \times 2.2$ arcmin^2). The last panel summarizes the evolution of the AMF for our nominal F200W filter.

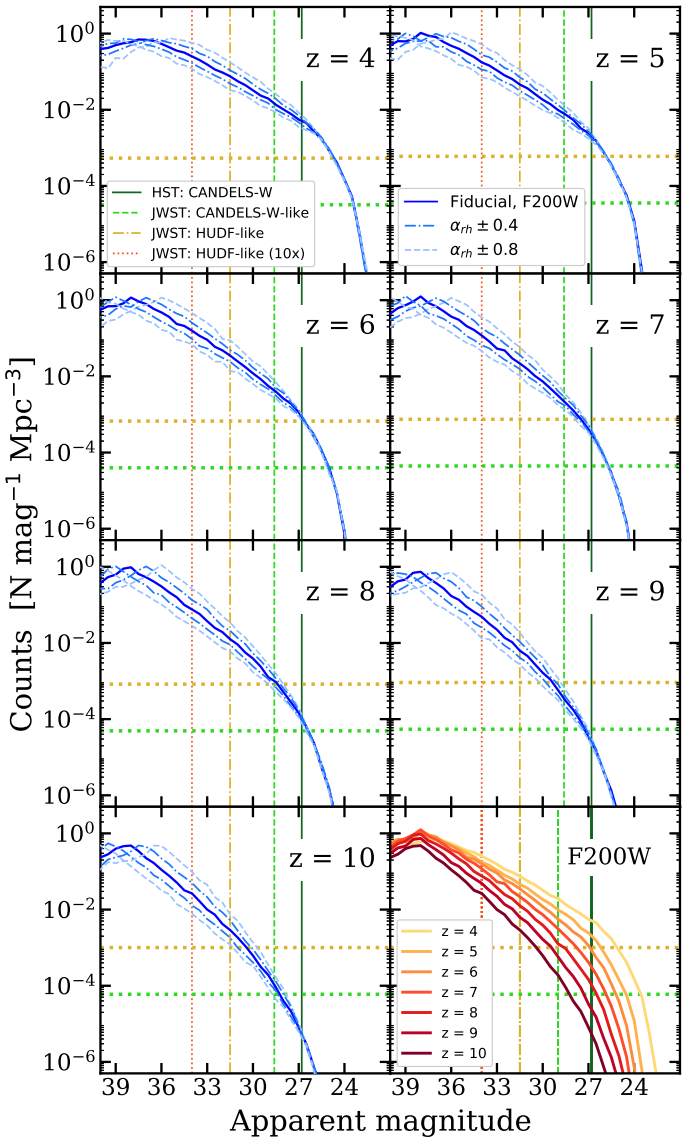

Figure 2.13 Redshift evolution of the AMFs in the F200W filter. The vertical dashed lines represent the detection limits for example *JWST* surveys similar to legacy *HST* counterparts; see Table 2.6 for details. The green and yellow horizontal dashed lines show where ten objects are expected in a ~ 100 arcmin^2 survey and in the HUDF field (2.4×2.4 arcmin^2), respectively. Blue dashed lines show the cases where we let $\alpha_{\rm rh} = 2.4$ (above) and 3.2 (below), and light blue dot-dashed lines show the cases where we let $\alpha_{\rm rh} = 2.0$ (above) and 3.6 (below). The last panel summarizes the evolution of the AMF for the F200W filter.

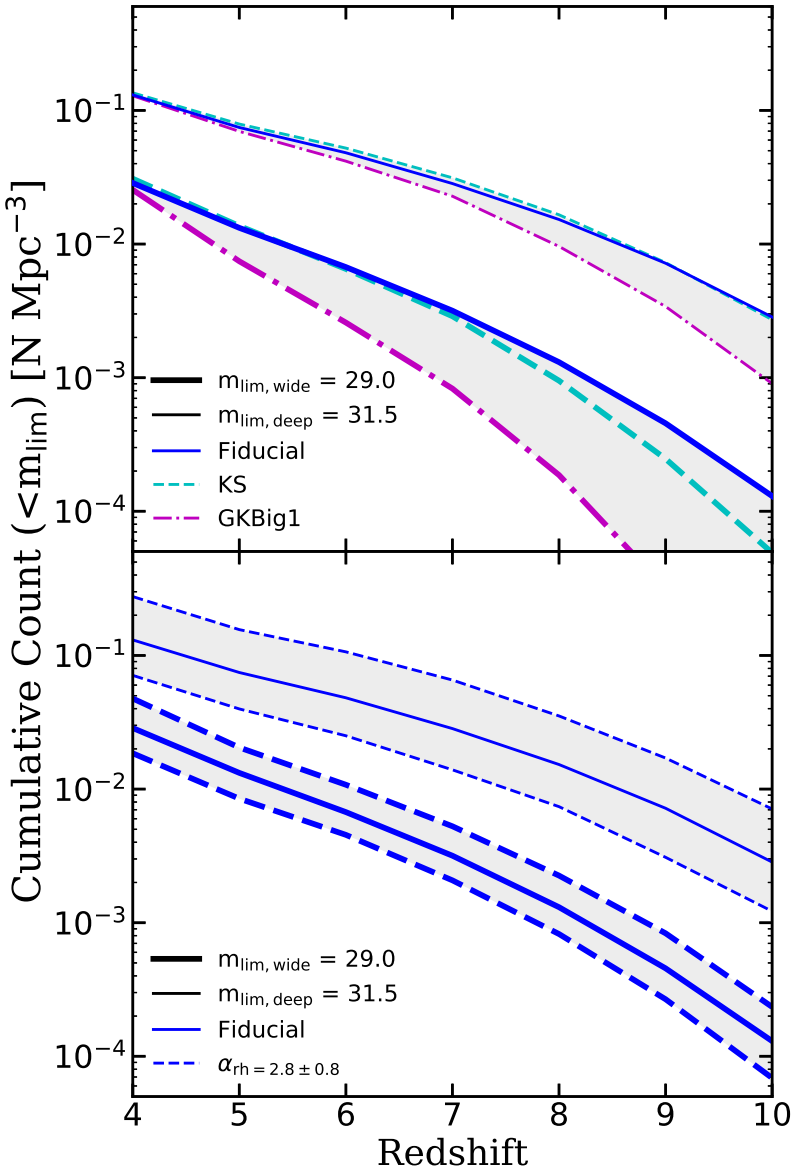

Figure 2.14 The number density of objects brighter than some specified apparent magnitudes in the NIRCam F200W filter, $m_{\rm lim}$, where the bold (narrow) line shows $m_{\rm lim,wide} = 29.0$ (31.5) that is chosen to be close to the detection limit of a *JWST* investment similar to CANDELS-W (HUDF). The blue solid and dashed lines in both panels are identical and they serve as a visual guide for the predictions from our fiducial model. In the *upper panel*, the purple dot-dashed line shows the GK-Big1 model and the cyan dashed line shows the KS model. In the *lower panel*, the cyan dashed line shows the results for $\alpha_{\rm rh} = 2.0$ and the purple dot-dashed line shows the results for $\alpha_{\rm rh} = 3.6$. Light gray bands are added to visually group the lines that share the same $m_{\rm lim}$.

Table 2.6 Detection limits for selected *HST* blank field surveys and for comparable anticipated *JWST* legacy surveys. *JWST* prediction limits are estimated assuming the use of the F200W filter. At the end we show a bracketing extreme case where *JWST* conducts a HUDF-like survey on a cluster field with 10x magnification.

	Survey Name	Detection Limit[a] [AB Mag]
HST	CANDELS - Wide	26.8
	CANDELS - Deep	27.8
	HUDF	29.5
	Frontier Field (unlensed)	29.0
	Frontier Field (10x mag.)	31.5
JWST	CANDELS-Wide-like	28.6
	HUDF-like	31.5
	HUDF-like (10x mag.)	34.0

Table 2.7 The best-fit Schechter parameters for the AMFs for the NIRCam F200W filter between $z = 4 - 10$ predicted by our fiducial model.

z	ϕ^* [10^{-3} Mpc^{-3}]	M^* [AB Mag]	α
4	2.423	24.748	-1.526
5	1.709	25.576	-1.589
6	0.989	25.955	-1.676
7	0.569	26.281	-1.733
8	0.222	26.427	-1.830
9	0.113	26.767	-1.881
10	0.052	27.163	-1.950

that shape galaxy properties at $z \gtrsim 6$; these include cosmological accretion, strong stellar-driven winds that are more efficient at low masses, and (more controversially), black hole feedback that preferentially suppresses star formation at high masses. However, these processes can sometimes have degenerate effects on galaxy properties. Although differentiating the effects of these processes is extremely challenging, being able to do so is crucial for interpreting galaxy observations at all redshifts. Observational constraints at $z \gtrsim 6$ are currently quite limited, leaving significant uncertainties in the current theories of galaxy formation and their predictions at very high redshifts. One of the main highlights of this work is that we explored how variations in several uncertain physical processes in theoretical models will impact the global physical properties of galaxies and what may be seen by future observations. We also discuss how these processes may be disentangled with the significant insights *JWST* observations have to offer.

High-z star-forming galaxies are most luminous in the observed frame infrared, making them observable with NIRCam and MIRI onboard *JWST*, and it has been shown that the choice of broadband filters can have an impact on the photometric redshift estimation and the follow-up interpretation (Bisigello et al. 2016, 2017). In Fig. 2.12, we demonstrated how objects drop out from

shorter wavelength filters starting at $z \sim 5$, making the F200W filter the nominal choice for blind surveys for these deep-field objects across $z = 4$–10 with investments comparable to CANDELS-Wide and HUDF. For galaxies beyond $z \sim 10$, observations in the longer wavelength filters, namely F356W and F444W, will be required with exposure times $\gtrsim 10^5$ seconds.

With its unprecedented sensitivity, *JWST* will be able to probe large populations of galaxies in deep space and provide constraints for the faint-end slope of the UV LFs at $z \lesssim 8$ and for the evolution of the bright-end up to $z \sim 10$. These observations will put the empirical relations derived from low-redshift observations to the test in extreme conditions. Moreover, the redundancy in multiple NIRCam broadband filters at lower redshifts will enable multi-wavelength measurements, which is crucial to recovering the physical properties of high-z galaxies and breaking degeneracies in the underlying physical processes.

In this work, we utilized a computationally efficient SAM to make quantitative estimates of observable properties for galaxy populations expected to be detectable with *JWST*. The physically motivated empirical recipes adopted in the model are well-tested at lower redshifts ($z \lesssim 6$, see SPT15). We present model predictions at redshifts where these models have never been tested before, and where observational constraints are relatively limited. By exploiting the efficiency of our model, we also provide forecasts for various model variations, including exploring multiple SF recipes and parametrizations of outflow rates for stellar-driven winds.

Physical processes that shape the formation of galaxies are degenerate, and yet each of them evolves slightly differently over time and can simultaneously affect multiple physical properties. Traditionally, there has been tension for galaxy formation models to simultaneously match the observed gas fraction, stellar metallicity, and stellar fraction. In order to calibrate our model, we carefully balance the model parameters for multiple physical processes to match observations at $z \sim 0$. Fig. B.1 summarizes how the outputs of our calibrated model compare with observational constraints. See appendix for details of the calibration process and tests.

Massive star forming galaxies in the lower redshift Universe are obscured by dust in the line of sight, which makes it extremely difficult to determine the true underlying stellar content. In our model, the attenuation effect due to dust in the galactic disc is estimated based on galaxy physical properties, incorporated with a dust-to-metal ratio guided by observations. In agreement with previous studies, we find that the optical depth of dust, or effectively the dust-to-metal ratio, is required to evolve with redshift. However, we also note that our very simple dust recipe does not accurately represent the complex geometries of dust relative to stars that may be common in high redshift galaxies (Koprowski et al. 2016; Chen et al. 2015).

We find that star formation remains fairly efficient in low-mass halos, in conflict with the

Krumholz & Dekel (2012) model, which argues that star formation is heavily suppressed at $z > 2$ in dark matter halos with masses $< 10^{11}$ $M_\odot$ due to the low metallicity of gas in these halos. However, recent lensed and very deep fields observations provide strong support for LF remaining fairly steep down to low halo. masses. We are aware of the fact that the underlying assumptions of these metallicity-based multiphase-gas partitioning recipes tend to break down in extremely low-metallicity environments (e.g. $Z_{\rm cold} < 0.05 Z_\odot$ (Krumholz et al. 2009; Gnedin & Kravtsov 2011), which might cause our model to over-predict the stellar content in our least massive halos and at higher redshifts.

Cosmic reionization in our models is treated with a rather ad-hoc prescription, in which the whole universe is reionized uniformly and instantaneously. Our models currently do not attempt to model the formation of metal-free Population III stars explicitly. Instead, we set a metallicity floor of $Z_{\rm pre\text{-}enrich}$ to represent pre-enrichment in the initial hot gas in halos and the gas accreted onto halos due to cosmological infall.

2.5.1 Physical processes that shape the bright end of UV LFs

The strong evolution in the bright end of the UV LFs indicates a rapid growth of massive galaxies in the early universe. However, the physical processes that drive or regulate this evolution are still unclear. Due to the limited direct observations at $z \gtrsim 6$, it is extremely difficult to disentangle the effects of multiple processes.

Each observed property of galaxies is resulting from a unique combination of underlying physical processes. Although the underlying fundamental physics should remain the same at all times, since the physical processes that drive the evolution of these objects depend on a number of effectively redshift dependent conditions, the global properties of galaxies may also be effectively redshift dependent. The evolution of these resulting global physical properties change subtly depending on the strength of the impact of each of the physical processes. Therefore, being able to disentangle the impact of these driving physical processes, as well as the evolution of the resultant observable properties, is most important for understanding galaxy formation and for making accurate predictions for the galaxies that are yet to be directly observed.

In this work, we have shown that the choice of star formation recipe and the gas depletion time both have a very strong impact on the abundance of bright galaxies. We have also provided model outputs where we systematically vary some of the physical recipes and associated parameters. For instance, changing the gas depletion timescale by a factor of two in either direction seems to mainly impact galaxies of $M_{\rm UV} \lesssim -19$. Under this variation, the bright end of the UV LFs deviates mildly from the fiducial results but still remains mostly within the uncertainties in the observations. On

the other hand, altering the SF recipes would significantly change the predicted number density for luminous galaxy populations, where the distinction occurs at $M_{\rm UV} \lesssim -14$ at $z = 10$, and this limit gradually evolves to $M_{\rm UV} \lesssim -20$ at $z = 4$. The abundance of faint galaxies is largely insensitive to either of these processes due to self-regulation of SF in low-mass halos, which will be discussed in detail in the next subsection.

To focus on the impact of these processes on the underlying stellar populations, in Fig. 2.2 we compare the intrinsic UV LFs in the absence of dust. The free parameters in the empirical SF recipes that are calibrated to $z \sim 0$ observations and the SF relation do not evolve over time, with the fiducial GK-Big2 and GK-Big1 representing the more and less optimistic SF scenarios, respectively. Both our fiducial model and the classic KS model are able to match high-z observations quite well without further parameter retuning. Given that the general consensus is that high-z galaxies are much less dusty than their low-z counterparts, these empirical SF recipes might hold to a certain extent even in extreme environments. Interestingly, GK-Big1 is able to match observations quite well at low redshifts ($z \lesssim 6$) when dust attenuation is omitted, but it underpredicts the number of bright galaxies at higher redshifts. This implies that if the dust attenuation at these redshifts is significantly lower than what we have assumed, a mildly evolving $\Sigma_{\rm SFR}$–$\Sigma_{\rm H_2}$ relation may be able to produce sufficient numbers of bright galaxies at high redshifts.

It is well known that at $z \sim 2$–4, the most rapidly star-forming galaxies are heavily obscured by dust. The dust content of galaxies at higher redshift is uncertain. Since dust is built up over generations of star formation, the general expectation is that high redshift galaxies should be less dusty than lower redshift ones (Popping et al. 2017b). Moreover, there is mounting evidence that the dust geometry may be different in high redshift galaxies, giving rise to differences in attenuation for a given dust mass (Popping et al. 2017a; Narayanan et al. 2018). Most theoretical models that include the effects of dust assume a relationship between dust optical depth and the metallicity and gas surface density, similar to the one we have adopted here. Our simple empirical model for dust is calibrated guided by available observations at $z = 4$–10, which requires a redshift dependent dust-to-metal ratio in order to be able to simultaneously match constraints at all redshifts within the framework of our simple model. Our results show that dust attenuation seems to be less effective at $M_{\rm UV} \gtrsim -19$, however, with a much stronger redshift evolution comparing to bracketing cases for gas depletion times we showed. Future multi-wavelength observations with multiple NIRCam filters may help constrain the dust content (e.g. using the UV continuum slope β as in Finkelstein et al. 2012b, see also Popping et al. 2017a).

Using similar SAMs and a similar dust model, S12 found that their models with a Calzetti attenuation curve underproduced UV-luminous galaxies at high redshift ($z \sim 3$–5). Our updated

models perform better in this regard, as a result of the updated Planck cosmology and associated recalibration. As shown in Fig. 2.1, the amount of dust required to match observations between = 4–6 has been reduced by roughly a factor of two. However, the UV LFs remain quite sensitive to $\tau_{\rm dust,0}$ and, as pointed out in previous works, using a fixed dust parameter that is normalised to observations at low redshifts would systematically underproduce UV-luminous galaxies at high redshift. In future works we plan to explore models with a more self-consistent treatment of dust formation and destruction and more complex treatment of dust geometry, and the implications for Far-IR and mm and sub-mm observations (Popping et al. in prep).

Although we find that black hole feedback plays a negligible role in shaping galaxy properties at these redshifts, this may be due to the specific manner in which black hole seeding, growth, and feedback are implemented in these models. AGN feedback is implemented in our models via two different modes: 'jet mode' (also called 'radio mode') and 'radiative mode' (also called 'quasar mode' or 'bright mode'). 'Radio mode' feedback (Croton et al. 2006, S08) is implemented as a heating term which (in the Santa Cruz SAMs) scales as a power-law function of the black hole mass, based on observations of radio jets in nearby galaxy groups and clusters (see S08). The jets are assumed to be able to couple efficiently with the hot gas only when the cooling time is longer than the dynamical time, which tends to be the case in massive halos at late times. As a result, jet mode feedback becomes effective only in massive halos at redshifts below about $z \sim 1$.

The Santa Cruz SAMs also include 'radiative mode' feedback, in which cold gas can be removed from galaxies via radiation pressure driven winds associated with radiatively efficient accretion onto a black hole. Although this mode can act at high redshift, it tends to have little effect on galaxy properties, because it is assumed that the winds can only act 'ejectively' on the cold interstellar gas, and new gas tends to cool rapidly and replenish the cold gas reservoir, especially at high redshift. However, the treatment of radiative mode feedback in the existing models is based on an older suite of hydrodynamic simulations of binary mergers between idealized galaxies with no initial hot gas halo. More recent cosmological zoom-in simulations including thermal and kinetic feedback from radiatively efficient black hole accretion find that there is also a strong *preventative* feedback effect, as these winds can significantly reduce the density of gas near the centres of halos and thereby suppress cooling for much longer timescales (Choi et al. 2015, 2017a; Brennan et al. 2018). In this picture, we might expect quenching via radiative mode feedback to be more effective at higher redshifts.

2.5.2 Physical processes that shape low-luminosity galaxies

In this work, we updated the Santa Cruz SAM treatment of photoionization squelching by implementing the fitting function presented in Okamoto et al. (2008) to model heating from an photoionizing background. Our results in Fig. 2.3 show that the effect is negligible at all redshifts. Some studies suggest that photoionization squelching plays a significant role in shaping the faint end of the galaxy LF, suppressing the faint end of the UV LFs in the range $M_{\rm UV} \sim -12$–-10 (Shapiro et al. 2004; Iliev et al. 2005; Ocvirk et al. 2016). Some others have shown that dense clumps are extremely hard to penetrate and become ionized due to a high recombination rate. Simulations have shown that cooling in gas clumps should be fairly efficient when densities are comparable to the virial density of a halo (Noh & McQuinn 2014). Susa & Umemura (2004a,b) used radiative hydrodynamic simulations with radiative transfer to show that a photoionizing background is devastating only for low-mass systems of $V_c \lesssim 20$ km s^{-1}, which are halos close to the atomic cooling limit. Similar results are also presented by Okamoto et al. (2008), where only extremely low density gas clumps are completely evaporated and the overall effect of a photoionizing background on galaxy formation is much weaker than previously thought.

In our models, stellar-driven winds play a dominant role in shaping the slope of the faint-end of the UVLF as well as the location of the turnover. This type of simple empirical model for stellar-wind feedback has been widely adopted in both SAMs and some cosmological numerical hydrodynamic simulations. The recipe has been shown to be quite successful in reproducing observations at lower redshifts for galaxies across a wide range of masses. As discussed in SPT15, SF in low-mass halos is self-regulated by stellar feedback. For instance, when SF becomes more efficient, more gas is ejected by energetic stellar winds, and hence reducing the supply of cold gas and yielding less efficient SF, and vice versa. This effect has also been examined and demonstrated in a number of studies (Schaye et al. 2010; Haas et al. 2013; White et al. 2015). Here we show that this remains the case in low-mass halos up to very high redshift, even in models with metallicity dependent, H_2-based star formation.

The free parameters in our physically motivated stellar feedback recipe have only been tuned to match observed SMF and stellar-to-halo mass ratio at $z \sim 0$. In addition to the calibrated fiducial values, we also show example cases where we systematically vary the SN feedback slope $\alpha_{\rm rh}$ slightly by ± 0.4 and ± 0.8 while keeping the rest of the model unchanged. Our results clearly demonstrate that this parameter indeed plays a dominant role even at extreme redshifts. However, we also show that the more massive galaxies are insensitive to variations in this parameter, because stellar-driven winds cannot efficiently escape the deep potential wells of the halos that host these objects.

Our model is one of the very few simulations that is capable of resolving objects as tiny as $V_c \sim 20$

$\mathrm{km\,s^{-1}}$in a cosmological context, close to the atomic cooling limit. We find that the turnover in the UVLF due to the atomic cooling limit occurs at around $M_{\mathrm{UV}} \sim -8$ at $z \sim 4$, moving slightly brighter to about $M_{\mathrm{UV}} \sim -9$ at $z \sim 10$. We find that the magnitude where the turnover occurs shifts brighter or fainter by up to ~ 1 magnitude under variations in the stellar feedback parameter α_{rh}. This is because α_{rh} changes the slope of the median relationship between halo mass and stellar mass or SFR, shifting the rest-UV magnitude corresponding to a halo circular velocity of $V_c \sim 20$ $\mathrm{km\,s^{-1}}$.

2.5.3 Probing ultra high-redshift galaxies beyond *JWST*

The galaxy populations predicted by our model span a wide range of luminosities and redshifts. Even though *JWST* will be able to detect many objects during the epoch of reionization, our models predict that there will still be a significant population of objects too faint to be detected even by *JWST*. Although these galaxies are unlikely to play a significant role in cosmic reionization, they are thought to have hosted Pop III stars and polluted the ISM with the first heavy elements. Therefore, constraints on these objects are a very important missing piece in the current formation theory for stars and galaxies. Probing these objects directly requires instruments with sensitivity many times higher than what *JWST* has achieved. However, the bulk effects of these objects can be studied via metal absorption (e.g. Finlator et al. 2013) or intensity mapping (Visbal & Loeb 2010; Visbal et al. 2011). Many ongoing and planned intensity mapping pathfinders (e.g. CHIME, HERA, HIRAX, Tianlai, BINGO, LOFAR, MeerKat, CONCERTO, STARFIRE) are paving the way to future high-z large-scale multiline intensity mapping surveys.

On the other hand, planned wide-field surveys, such as those that will be carried out with Euclid and the Wide-Field Infrared Survey Telescope (WFIRST), will probe unprecedented areas, providing constraints on the massive, bright galaxy populations (Racca et al. 2016; Spergel et al. 2015).

We plan to exploit our model framework to forecast and provide an interpretive framework for these and other observations in future projects. We plan to further explore the progression of cosmic reionization arising from our predicted galaxy populations, and compare that to the latest observational constraints from the Lyα forest and the Thomson scattering optical depth for the cosmic microwave background (Yung et al. 2020a).

2.6 Summary and Conclusions

In this work, we presented predictions for galaxy populations that are expected to be detected in upcoming *JWST* observations, and showed how they can potentially constrain the physical processes that govern the formation and evolution of these objects. Our galaxies are modelled using the well-established Santa Cruz semi-analytic model with the recently updated multiphase gas partitioning and H_2-based SF recipes. We also used semi-analytic dark matter halo merger trees that are constructed based on the EPS formalism to achieve the very wide dynamic range and computational efficiency required for forecasting observable properties of galaxy populations over a wide range of masses and redshifts. Moreover, we adopted the updated Planck cosmology and the model parameters were recalibrated to match the latest observational constraints near $z \sim 0$.

By exploiting the high efficiency of our model, we were also able to systematically vary the SF recipes, as well as the sub-grid physical parameters for gas depletion timescale and stellar feedback relation slope, in a controlled manner. We use these results to explore the physical processes that have degenerate effects on galaxies, which create tension in matching galaxy properties and other cosmological observables. We also discuss whether these processes can be disentangled and what we expect to learn from the upcoming deep-field observations.

Predictions for rest-frame UV luminosity functions at $z = 4$–10 are presented and are compared to existing observations and other models. We include the effects of dust attenuation using an empirical dust recipe with a redshift dependent dust-to-metal ratio. Although the free parameters in our model are only calibrated to match observations at $z \sim 0$, our results matches surprisingly well with UV LF constraints at $z > 4$. In addition, our results agree extremely well with previous theoretical studies, particularly for more luminous galaxies. We predicted that the faint end of the UV luminosity functions will remain steep below the current detection limit until $M_{\rm UV} \sim -9$. We showed that the gas depletion time and the choice of star formation recipe have strong influences on star formation in luminous galaxies. Conversely, the effect of feedback from AGN is found to be negligible, although this may be due to shortcomings in our modelling of black hole seeding, accretion, and/or feedback. However, starlight from the most intrinsically UV-luminous galaxies is also heavily obscured by dust, making it extremely difficult to disentangle these degenerate effects from multiple physical processes. On the other hand, star formation in low-mass halos seems to be most strongly affected by stellar feedback, with photoionization feedback having a negligible effect on the populations that we studied.

We estimated the apparent magnitudes of our predicted galaxy population utilizing the published *JWST* NIRCam broadband filters and presented them in the form of one-point distribution

functions, which may be compared directly with observations. We also estimated the effects of dust attenuation and illustrated the sensitivity to stellar feedback efficiency. We show that, with a simple dust model, the effect of dust is only significant for rapidly star forming, metal rich galaxies and is more important at lower redshifts.

We summarize our main conclusions below.

1. A relatively simple and computationally efficient semi-analytic model, which has been calibrated only to $z \sim 0$ observations, produces predictions that agree remarkably well with observed UV luminosity functions from $z \sim 4$–10.

2. Star formation physics and gas depletion time are dominant in determining the abundance of bright, massive galaxy populations, and these physical processes are degenerate with the attenuation due to dust. Therefore, it is critical to obtain independent probes on the dust content of high redshift galaxies in order to be able to disentangle the underlying physics.

3. The faint-end slope of UV LFs is mainly sensitive to the scaling of the mass-loading factor for stellar driven winds with halo or galaxy properties. We find that the effect of photoionization squelching on galaxies that will be detectable with *JWST* is negligible. *JWST* observations will be able to place important constraints on stellar feedback. However, *JWST* will not be able to probe down to the atomic cooling limit.

4. In our models, the absolute magnitude at which the 'turnover' in the UV LF occurs due to the atomic cooling limit is also sensitive to the adopted recipe for stellar feedback. This is because stellar feedback changes the halo mass that hosts galaxies of a given UV luminosity.

Chapter 3
Physical Properties and Scaling Relations

The long anticipated *James Webb Space Telescope* (*JWST*) will be able to directly detect large samples of galaxies at very high redshift. Using the well-established, computationally efficient Santa Cruz semi-analytic model, with recently implemented multiphase gas partitioning and H_2-based star formation recipes, we make predictions for a wide variety of galaxy properties for galaxy populations at $z = 4$–10. In this work, we provide forecasts for the physical properties of high-redshift galaxies and links to their photometric properties. With physical parameters calibrated only to $z \sim 0$ observations, our model predictions are in good agreement with current observational constraints on stellar mass and star formation rate distribution functions up to $z \sim 8$. We also provide predictions representing wide, deep, and lensed *JWST* survey configurations. We study the redshift evolution of key galaxy properties and the scaling relations among them. Taking advantage of our models' high computational efficiency, we study the impact of systematically varying the model parameters.

This chapter is reproduced from published work *Semi-analytic forecasts for JWST – II. Physical properties and scaling relations for galaxies at* $z = 4$–10 (Yung et al. 2019. MNRAS, 490, 2855).

3.1 Introduction

The highly anticipated *James Webb Space Telescope* (*JWST*) will be equipped with extremely sensitive instruments that will be uniquely capable of detecting extremely distant faint galaxies. As postulated by the hierarchical structure formation paradigm within the ΛCDM (cosmological constant and cold dark matter) cosmological framework (White & Rees 1978; Blumenthal et al. 1984), low-mass objects are expected to be fairly abundant throughout the Universe, even at high redshifts. However, their properties and abundances remain largely unconstrained since direct detections for these objects with current instruments are not possible.

In contemporary deep-field astronomical surveys, galaxy candidates at $z \gtrsim 2$ have been routinely identified using the 'Lyman-break' photometric selection technique (Steidel & Hamilton 1992, 1993; Steidel et al. 1996). This selection is carried out by using a set of strategically designed filters to identify the occurrence of the redshifted Lyman-limit discontinuity, caused by an intrinsic spectral

break combined with absorption by intergalactic neutral hydrogen along the line of sight. Over the past decade, space-based surveys, such as the *Hubble* Ultra Deep Field (HUDF; Beckwith et al. 2006, see also Bouwens et al. 2011; Ellis et al. 2013; Oesch et al. 2013) and the Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), and ground-based surveys, such as the United Kingdom Infrared Telescope Deep Sky Survey (UKIDSS; Warren et al. 2007) and UltraVISTA (McCracken et al. 2012), have found nearly 2000 galaxy candidates at $z = 6 - 10$ using this technique, with faint objects reaching absolute UV magnitude ~ -17 at $z \sim 6$. Observations can reach an even fainter detection limit in fields that are gravitationally lensed by massive foreground clusters (Castellano et al. 2016; Laporte et al. 2016; Kawamata et al. 2016; Bouwens et al. 2017; Livermore et al. 2017; Lotz et al. 2017; Atek et al. 2018; Ishigaki et al. 2018). However, the uncertainties from magnification and foreground cluster modelling associated with these detections are relatively large (Bouwens et al. 2017).

The Near-Infrared Camera (NIRCam), one of *JWST*'s onboard photometric instruments, possesses unprecedented infrared (IR) sensitivity, which is expected to enable the detection of distant galaxies far below the detection limits of past or current facilities. *JWST*'s spectrometers, the Near-Infrared Spectrometer (NIRSpec) and Mid-Infrared Instrument (MIRI), will be able to provide follow-up spectroscopic studies for more luminous high-redshift galaxy candidates. A sizable amount of *JWST* observing time has already been allocated, dedicated to the search for high-redshift galaxies. At the beginning of its mission lifespan, *JWST* is expected to carry out a number of Guaranteed Time Observation (GTO) and Early Release Science (ERS) projects that are aimed at studying galaxies and the intergalactic medium at high redshift, such as the *JWST* Advanced Deep Extragalactic Survey (JADES; Williams et al. 2018), the Cosmic Evolution Early Release Science survey (CEERS; Finkelstein et al. 2017), Grism Lens-Amplified Survey from Space (GLASS; Treu et al. 2015), and Targeting Extremely Magnified Panchromatic Lensed Arcs and Their Extended Star Formation (TEMPLATES; Rigby et al. 2017). Aside from these, there are also ongoing projects that are making use of other facilities to prepare for *JWST*, such as the Magellan Evolution of Galaxies Spectroscopic and Ultraviolet Reference Atlas project (MEGaSaURA; Rigby et al. 2018).

A fair amount of effort has been dedicated to developing models to connect the observed, photometric properties to the underlying, inferred physical properties of these galaxies, as well as the physical processes that drive their formation. Physical properties that have been estimated directly based on observations include star formation rate (SFR), stellar mass, stellar age, metallicity, and radial size. Since the baryonic fraction and temperature of the Universe affect the properties of the first galaxies and the first stars, the physical properties of these emerging galaxy populations in the early universe are imprinted with the conditions of the cosmic environment at the time they were

formed. Therefore, these galaxies also serve as indirect probes of the underlying cosmology, especially the overall matter density and the baryonic fraction. Moreover, there are collective properties that can be measured, such as the cosmic star formation rate density and global stellar mass density (e.g. Madau & Dickinson 2014).

Furthermore, combining more than one observed or inferred property can reveal a scaling relation between the two. These scaling relations among galaxy properties have been studied extensively for decades and yield fundamental insights about the physics of galaxy formation. Scaling relations may cover a large variety of properties in addition to the ones mentioned above. Some well-known examples are the Faber-Jackson relation (Faber & Jackson 1976), the Tully-Fisher relation (Tully & Fisher 1977), the Kennicutt-Schmidt relation (Kennicutt 1989), the mass-metallicity (MZR) relation (McClure & van den Bergh 1968; Lequeux et al. 1979; Tremonti et al. 2004; Gallazzi et al. 2005; Zahid et al. 2013), and the stellar mass-SFR relation, sometimes called the star formation main sequence (e.g. Brinchmann & Ellis 2000; Brinchmann et al. 2004; Noeske et al. 2007; Wuyts et al. 2011). We may also consider scaling relations between quantities predicted by theory that cannot be directly measured from observations, such as the stellar mass vs. halo mass relation or galaxy size vs. halo size relation (Moster et al. 2010, 2013, Behroozi, Wechsler, & Conroy 2013, Somerville et al. 2018). These relations are collectively affected by many physical processes, both local (ISM scale) and global (galaxy scale), which provide extremely important insights regarding the formation and assembly histories of these objects. On the other hand, the intrinsic scatters in these relations also hint at whether there are important higher order parameters (i.e. whether the scaling relation actually sits in a higher dimensional space, such as fundamental plane relations). Next-generation observing facilities will constrain the evolution of these scaling relations over cosmic time.

One of the main science goals of *JWST* is to constrain the nature of the sources that reionized the Universe. Recent studies have shown that high-redshift low-mass galaxies could have been the major source of the ionizing photons that reionized the Universe (Kuhlen & Faucher-Giguère 2012; Anderson et al. 2017; Finkelstein et al. 2019). However, there has historically been tension between the efficient SF in low-mass halos needed to reionize the Universe early enough to satisfy observational constraints, and the inefficient SF observed in low-mass galaxies today (Madau et al. 2008; Finkelstein et al. 2015a; Robertson et al. 2015).

Models and theoretical simulations set within the framework of cosmological structure formation are a powerful tool for creating forecasts for future observations. Inevitably, there are trade-offs between the breadth of physical processes that can be included, the accuracy with which these processes can be modelled, and computational limitations. A variety of different methods have been developed and employed to make predictions for galaxy populations that lie outside of the scope

of current observations. These include (sub)-halo abundance matching models, (semi-)empirical models, semi-analytic models, and numerical methods. There are advantages and disadvantages to each method, and on the whole they represent a complementary toolkit in this challenging landscape.

There has been a long standing history of using numerical methods to carry out *a priori* galaxy formation simulations, in which one attempts as much as possible to simulate the main physical processes explicitly. However, in practice, since these physical processes that influence galaxy formation operate over an extremely broad range of spatial and temporal scales, one must be selective about the physical prescriptions and resolution. Inevitably, in large-scale simulations, some 'sub-grid' recipes are needed to represent processes occurring at scales well below the physical resolution (see Somerville & Davé 2015 for a detailed discussion). In addition, modelling galaxy formation in a cosmological context is very challenging, given the tension between simulated volume and mass and spatial resolution. Moreover, due to computational limitations, numerical simulations focussed on the high-redshift Universe are often halted at some intermediate ($z \gtrsim 6$) redshift (e.g. O'Shea et al. 2015; Wilkins et al. 2017; Jaacks, Finkelstein, & Bromm 2019) and thus cannot be validated with low-redshift observational constraints. Zoom-in simulations are designed to allow high-resolution simulations to be conducted within a proper cosmological context, and are carried out by simulating a large cosmological volume at a coarse resolution and selecting and 'resimulating' a desired smaller region with higher resolution and additional prescriptions for baryonic physics (e.g. Hopkins et al. 2014, 2018; O'Shea et al. 2015). However, these techniques do not by themselves yield predictions for statistical properties of populations, as the resimulated halos are in general not representative of the underlying full cosmological distribution.

At the other end of the spectrum, purely empirical models extrapolate observed galaxy populations to higher redshifts and/or lower luminosities with no theoretical underpinning and no cosmological framework (e.g. Kuhlen & Faucher-Giguère 2012; Williams et al. 2018). Halo abundance matching models (and variants sometimes called semi-empirical models) attempt to derive relationships between dark matter halo properties and galaxy observables using lower redshift observations, then use the predicted evolution of halo properties from a cosmological model to make predictions for galaxy properties (Trenti et al. 2010, Trenti, Perna, & Jimenez 2015; Tacchella et al. 2013; Behroozi & Silk 2015; Behroozi et al. 2019; Mason, Trenti, & Treu 2015; Moster, Naab, & White 2018; Tacchella et al. 2018; Wechsler & Tinker 2018). Naturally, this requires assumptions about how the galaxy-halo relationship evolves in this unexplored territory, but these approaches are highly flexible and computationally efficient.

The semi-analytic modelling approach provides an attractive 'middle way' for modelling large populations of galaxies and exploring a large dynamical range in halo mass and environment, and

has been used for decades to make predictions for the properties of high-redshift galaxies (e.g. Somerville, Primack, & Faber 2001; Somerville et al. 2015; Henriques et al. 2015; Lacey et al. 2016; Poole et al. 2016; Rodrigues, Vernon, & Bower 2017). This approach is built within the framework of cosmological structure formation, and adopts parametrized phenomenological recipes for physical processes such as cosmological accretion and cooling, star formation and black hole growth, and feedback from massive stars, supernovae, and AGNs. The parameters in these recipes are typically calibrated to reproduce key observational relationships for nearby galaxies. An advantage over the semi-empirical approach is that the parameters represent *physical quantities* that can frequently be constrained via observations, and compared with similar quantities in fully numerical hydrodynamical simulations. At the same time, the approach is still highly computationally efficient, allowing exploration of parameter space and different physical recipes.

In this thesis chapter, we make predictions for galaxy populations at $z = 4$–10 with forecasts tailored specifically for upcoming *JWST* observations, and we investigate how uncertainties in the physical processes may affect the global properties of these galaxies. The key components of this chapter are summarized as follows: the semi-analytic framework used in this work is summarized briefly in Section 3.2. We present the physical properties and scaling relations for these galaxy populations in Section 3.3–3.5. We then discuss our findings in Section 4.4, and summary and conclusions follow in Section 3.7.

3.2 The Semi-Analytic Framework

The Santa Cruz semi-analytic model used in this work is slightly modified from the one outlined in Somerville, Popping, & Trager (2015, hereafter SPT15). We have implemented the Okamoto, Gao, & Theuns (2008) photoionization feedback recipe and updated the cosmological parameters to be consistent with the ones reported by the Planck Collaboration in 2015. The model components that are essential to this work have been concisely summarized in Chapter 2 and we refer the reader to the following works for full details of the modelling framework: Somerville & Primack (1999); Somerville, Primack, & Faber (2001); Somerville et al. (2008, 2012); Popping, Somerville, & Trager (2014, hereafter PST14) and SPT15. The cosmological parameters adopted in this work are the following: $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8\,{\rm km\,s^{-1}Mpc^{-1}}$, $\sigma_8 = 0.831$, and $n_s = 0.9665$.

Dark matter halo merger histories, also commonly referred to as merger trees, are the backbone of our semi-analytic models for galaxy formation. In order to efficiently sample halos over a wide mass range, including halos from close to the atomic cooling limit to the rarest, massive objects expected to be found in high-redshift surveys, we adopted a merger tree algorithm based on the

Extended Press-Schechter (EPS) formalism (Press & Schechter 1974; Lacey & Cole 1993). These semi-analytic merger histories have been shown to be qualitatively similar to the ones extracted from N-body simulations (Somerville & Kolatt 1999; Somerville et al. 2008; Zhang et al. 2008; Jiang & van den Bosch 2014). At each output redshift, we set up a grid of root halos spanning the range in virial velocity $V_{\rm vir} \approx 20$–$500\ \mathrm{km\,s^{-1}}$, and assign their expected volume-averaged abundances based on the halo mass function from the Bolshoi-Planck simulation from the MultiDark suite (Klypin et al. 2016) with fitting functions provided in Rodríguez-Puebla et al. (2016). For each root halo in the grid, one hundred Monte Carlo realizations of the merger histories are generated, each traced down to progenitors of a minimum resolution mass of either $M_{\rm res} \sim 10^{10}\mathrm{M}_\odot$ or 1/100th of the root halo mass, whichever is smaller.

As implemented in the latest iteration of the model (PST14; SPT15), the disc component of each galaxy is divided into annuli and the cold gas content in each annulus is partitioned into an atomic (HI), ionized (HII), and molecular (H_2) component. Among the models implemented and tested in PST14 and SPT15, the metallicity-based, UV-background-dependent recipe, which is based on simulations by Gnedin & Kravtsov (2011, hereafter GK), yields the best results and hence was adopted as the fiducial model for multiphase gas partitioning. With the estimated surface density of molecular hydrogen ($\Sigma_{\rm H_2}$), observationally motivated, empirical H_2-based SF relations are used to model the surface density of SFR ($\Sigma_{\rm SFR}$) by Bigiel et al. (2008, hereafter Big; see also Wong & Blitz 2002, Bigiel et al. 2011, Leroy et al. 2011). Recently, evidence from both theory and observation suggests that the SF relation slope may steepen to ~ 2 at higher gas surface densities (Sharon et al. 2013; Rawle et al. 2014; Hodge et al. 2015; Tacconi et al. 2018). Hence, we have adopted a 'two slope' relation where star formation efficiency increases with increasing H_2 density (labelled as Big2), and a 'single slope' relation where star formation efficiency remains linearly related to H_2 surface density (labelled as Big1; see Fig. 1 and Eqn. 6 in SPT15). In addition, we included results from using the widely adopted 'classic' cold gas-based Kennicutt-Schmidt (KS) SF recipe in our comparison (Schmidt 1959, 1963; Kennicutt 1989, 1998).

Our model uses a set of physically motivated, phenomenological and empirical recipes to track the evolution of a wide range of global physical properties of galaxies. These standard recipes include cosmological accretion and cooling, stellar-driven winds, chemical evolution, black hole growth and feedback, and galaxy mergers. A partial list of physical properties tracked in our model include stellar mass, SFR, masses of multiple species of gas (ionized, atomic, molecular), and stellar and gas phase metallicity. The star formation and chemical evolution histories are combined with stellar population models and a simple prescription for dust attenuation to compute predictions of observable properties (rest- and observed-frame luminosity in any desired filter). Throughout this work, we use the SSP

models of Bruzual & Charlot (2003) with the Padova1994 (Bertelli et al. 1994) isochrones and assume a universal Chabrier stellar initial mass function (IMF; Chabrier 2003a).

These models have been extensively tested at lower redshifts ($z \lesssim 6$) in previous works. For instance, SPT15 and PST14 present results from $z \sim 0$–6 and showed that the model predicted physical properties generally agree with observations; these properties includes SFR, specific SFR, atomic and molecular gas density, stellar and cold gas metallicity, stellar mass function, stellar-to-halo mass ratio. Free parameters in our models are calibrated to a subset of $z \sim 0$ observations, including stellar-to-halo mass ratio, stellar mass function, stellar mass-metallicity relation, cold gas fraction versus stellar mass relation for disc-dominated galaxies, and the black hole mass vs. bulge mass relation. (see Appendix in Chapter 2 for details). Without retuning these parameters to match observational constraints at higher redshifts, Chapter 2 has shown that the predicted UV luminosity functions and the cosmic star formation rate (CSFR) agree well with observational constraints up to $z \sim 8$.

In Chapter 2, we found that the key model parameters that have strong effects on the predicted rest-frame UV luminosity are the SF time-scale ($\tau_{*,0}$), which effectively characterizes the gas depletion time as $\Sigma_{\rm SFR} \propto \tau_{*,0}^{-1}$, and stellar feedback relation slope ($\alpha_{\rm rh}$), which characterizes the dependence of the mass loading factor of cold gas ejected by stellar feedback on halo circular velocity:

$$\eta_{\rm out} = \dot{m}_{\rm out}/\dot{m}_* = \epsilon_{\rm SN}\,(V_0/V_{\rm c})^{\alpha_{\rm rh}} \tag{3.1}$$

where $\dot{m}_{\rm out}$ is the rate at which cold gas is ejected from the ISM by stellar feedback, $\dot{m}_*$ is the star formation rate, $V_{\rm c}$ is the circular velocity of the galaxy, normalised by an arbitrary constant $V_0 = 200\,{\rm km\,s^{-1}}$, and $\epsilon_{\rm SN}$ and $\alpha_{\rm rh}$ are tunable free parameters referred to as the SN feedback efficiency and SN feedback slope. In this chapter, we similarly explore the sensitivity of physical parameters such as stellar mass or SFR to varying these model parameters. See Chapter 2 and SPT15 for descriptions of model components that are not varied in this work, such as photoionization feedback, AGN feedback, disc sizes, stellar population synthesis, dust attenuation, chemical evolution, and the calibration process.

3.3 Physical Properties of High-Redshift Galaxies

One-point distribution functions of observable quantities or inferred physical properties are perhaps the most basic way to summarize the statistical characteristics of large populations of galaxies. These quantities have been probed independently with various tracers and observational constraints are fairly abundant. Furthermore, two-dimensional distributions for these quantities reveal how one

quantity scales with another. Moreover, additional clues for characterizing the physical processes and for disentangling the degeneracies among processes can be obtained via the redshift evolution of these distribution functions and scaling relations.

One of the main goals of this work is to connect observable quantities to the underlying physical properties for very high redshift galaxies. In this section, we show distribution functions for selected physical properties and scaling relations among these properties for galaxies in halos with masses ranging from $M_{\rm H} \sim 10^8$–10^{13} $\rm M_\odot$ at $z = 4$–10. We also quantify the impact of uncertainties in our physical recipes on the resultant galaxy properties. All binned distribution functions presented in this work are available for download online at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/.

This section first presents a series of predicted distribution functions for key physical properties such as stellar mass (M_*) and star formation rate (SFR) tailored to specific types of *JWST* surveys (Section 3.3.1). We then present results of various star formation models and of systematically varying the parameters characterizing several key physical processes, and examine their impact on the predicted galaxy populations (Section 3.3.2). We also present a comprehensive comparison to predictions from other models in the literature (Section 3.3.3), and show predictions for cold gas mass ($M_{\rm cold}$) and molecular gas mass ($M_{\rm H_2}$) (Section 3.3.4). Stellar masses and SFR are often inferred from galaxy broad-band photometry or nebular emission lines, while the cold gas content may be estimated from CO or dust continuum emission. There are significant uncertainties and inherent assumptions in estimating these physical properties from observables, and therefore it is interesting to confront these relationships with the forward modelling predictions from theoretical models.

3.3.1 Distribution functions in mock *JWST* surveys

Given the rather good agreement between our fiducial models and existing observations (shown in Chapter 2), we can use our models to forecast the physical parameters of the populations that we expect *JWST* to be able to study. In Chapter 2, we explored the detection of high-redshift galaxies with the *JWST* NIRCam broad-band filters and found that the F200W filter yields the highest number of detections across $z = 4$–10. Using the observed-frame luminosity calculated for this filter, we can easily estimate which galaxies will be detectable for a survey of a given sensitivity and area. In the results presented here, we do not include noise or other instrumental effects. We plan on including these in a future work.

As illustrated in Chapter 2, the unprecedented sensitivity of *JWST* will be able to probe objects that are several magnitudes fainter than the current detection limit of *HST* deep surveys. These

objects correspond to objects with stellar masses and SFRs several orders of magnitude lower than the current limits. In this section, we perform selections on the galaxy populations predicted by our fiducial model using the observed-frame IR luminosity calculated for the NIRCam F200W filter $m_{\rm F200W}$. See Fig. 13 in Chapter 2, for predicted distribution functions for observed-frame IR magnitude calculated for the F200W filter and cut-offs for the detection limits and survey area. Our calculations have taken the absorption by the intervening IGM into account. We consider three distinct observing scenarios, including representative wide, deep, and lensed surveys. The assumed depths and areas for these hypothetical surveys are summarized in Table 3.1, where the survey areas are chosen to be similar to their legacy *HST* counterparts and the depths are estimated assuming the use of the F200W filter. We assumed an average 10x magnification for a lensed survey on a cluster field, and since the survey volume is inversely proportional to the magnification, we simply assume the survey area is a 10th of what we assumed for a deep survey. For each assumed survey area, adopting redshift slices with a width $dz = 1$, we also work out the critical stellar mass above which we would expect to detect less than one galaxy in the probed volume. In reality, of course, the upper limit on the stellar mass that will be robustly probed by *JWST* will be strongly affected by field-to-field variance caused by galaxy clustering and the underlying large-scale structure, and this will depend weakly on the precise field geometry. We plan to make detailed predictions for cosmic variance in *JWST* fields in future works, but do not address this here.

Table 3.1 Summary of assumed detection limits for the NIRCam F200W filter and survey areas for representative *JWST* surveys.

Survey Type	Detection Limit	Survey Area
Wide-field	28.6	~ 100 arcmin^2
Deep-field	31.5	2×2.2^2 arcmin^2
Lensed-field	34.0	$\frac{1}{10}(2 \times 2.2^2)$ arcmin^2

Fig. 3.1 shows the fraction of galaxies expected to be detected in wide, deep, and lensed *JWST* surveys as a function of stellar mass, using a sliding boxcar filter of width $\Delta M_* = 0.2$ in stellar mass. Stellar masses corresponding to 50 and 90% completeness for each redshift and survey configuration can be read off from these plots. Note that the reason that the completeness does not reach a perfect value of unity even for rather massive galaxies is that some of these massive galaxies are predicted to be significantly attenuated by dust in our models, resulting in a large scatter in $m_{\rm F200W}$ and M_* (see Fig. 3.16–3.20).

Fig. 3.2 shows the 'observable' SMFs for the three survey configurations at $z = 4$–10. We also include the full range of predictions and observational constraints from Duncan et al. (2014) and Song et al. (2016) to guide the eye. We show the stellar mass where the SMF becomes 50%

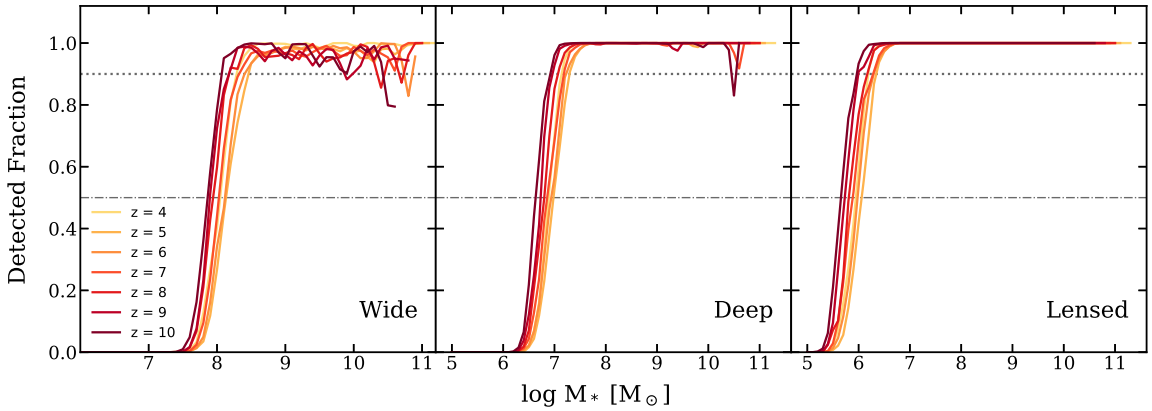

Figure 3.1 Fraction of galaxies detectable in wide, deep, and lensed *JWST* surveys from $z = 4$–10. The dotted and dot-dashed lines show detection fractions of 90% and 50%.

incomplete with vertical lines, and the number density where the expected number of objects in the survey drops below ten objects by horizontal lines. Thus, for a given survey area and depth, one can determine where the stellar mass function will be robustly probed by *JWST* by considering the part of the function that lies above the horizontal line and to the right of the vertical line of a given colour. Similarly, Fig. 3.3 shows the 'observable' SFRFs. We include the full range of observational constraints from Katsianis et al. (2017b), Katsianis et al. (2017a), and Smit et al. (2012) to guide the eye.

Although the populations that will be detected in wide-field NIRCam surveys seem to be fairly comparable to those of existing *HST* deep surveys (e.g. the CANDELS Deep survey has a H-band detection limit of $m_{\rm AB} \sim 27.8$ (Grogin et al. 2011), see also Table 2.6 in Chapter 2), current stellar mass and photometric redshift estimates rely heavily on observations from the Spitzer Space Telescope. *JWST* is expected to obtain much more precise photometry in the redder bands, which will also improve the photometric redshift measurements, as well as physical parameter estimates. Moreover, spectroscopic detections at $z \sim 4$–6 from NIRSpec and MIRI will provide additional constraints for these populations. It is also quite encouraging to see that lensed surveys will bring significant improvements to detecting low-stellar-mass galaxies, similar to the ultra-faint dwarf galaxies found in the local group. Our predictions also show crudely where we can expect *JWST*'s very limited field of view to cause the errorbars on the abundances of massive galaxies to become very large due to poor sampling and field-to-field variance.

In the bottom right panel of Fig. 3.2 and 3.3, we illustrate the redshift evolution of the populations expected to be detectable in a deep-field survey. It is noteworthy that the expected stellar mass and SFR corresponding to a given completeness limit evolves rather little from $z \sim 4$–10, because our models predict that high-redshift galaxies are intrinsically brighter than their low-redshift

counterparts of similar masses due to their overall younger stellar populations and higher SFR. In addition to that, massive galaxies at high redshifts have higher dust-extincted luminosities due to the lower dust content. This will be briefly discussed in §5 and further investigated in great detail in Chapter 4.

3.3.2 Evolution of stellar mass and star formation rate distributions for galaxy populations

Fig. 3.4 shows the redshift evolution of stellar mass functions (SMFs) between $z = 4$–10 predicted by the three different SF models; GK-Big2 (fiducial), GK-Big1, and KS. As shown in SPT15, all of these models produce results that qualitatively agree with observations at $z = 0$. However, due to their differences in gas depletion time (defined as the molecular gas mass divided by the star formation rate), the predictions from these models can vary quite a lot at high redshift. As shown in Fig. 14 of SPT15, the gas depletion time is shorter in massive galaxies at high redshift in the GK-Big2 model because of the steeper slope of the relationship between Σ_{H_2} and Σ_{SFR} in dense gas (see Fig. 1 and Eqn. 6 of SPT15).

The low-mass end of the SMFs is quite insensitive to changing the star formation recipe, while the abundance of massive galaxies at high redshift becomes very sensitive to the scaling of the H_2 consumption time. The reason for this is that star formation is regulated by stellar feedback in lower mass galaxies and at lower redshifts, while in massive galaxies at high redshift, the time to convert H_2 into stars becomes the rate limiting factor in star formation. Similar behaviour was seen in the response of the rest-UV LF to changes in the star formation recipe in Chapter 2. See SPT15 and SD15 for a more detailed discussion.

Our predictions are compared to observational constraints from Duncan et al. (2014) and Song et al. (2016). It is quite encouraging that even though our model is only calibrated to $z \sim 0$ observations, predictions from our fiducial model at these high redshifts still show overall very good agreement with these observational estimates. However, at $z \sim 4$, the low-mass-end slope is significantly steeper than that reported by Song et al. and seems to favor the Duncan et al. observational estimates. Note that the difference between our fiducial (GK-Big2) and GK-Big1 models is the Σ_{SFR}-Σ_{H_2} relation adopted in the SF recipe, which in the former case steepens in H_2 dense regime. The steepening feature is crucial for reproducing the current observational constraints.

The Duncan et al. (2014) and Song et al. (2016) observational estimates of the stellar mass function show significant discrepancies with one another, particularly at the high-mass end at $z \sim 7$ and the low-mass end at $z \sim 4$. The Duncan et al. study is based on the CANDELS GOODS South field, and Song et al. is based on observations from the CANDELS GOODS fields, the HUDF,

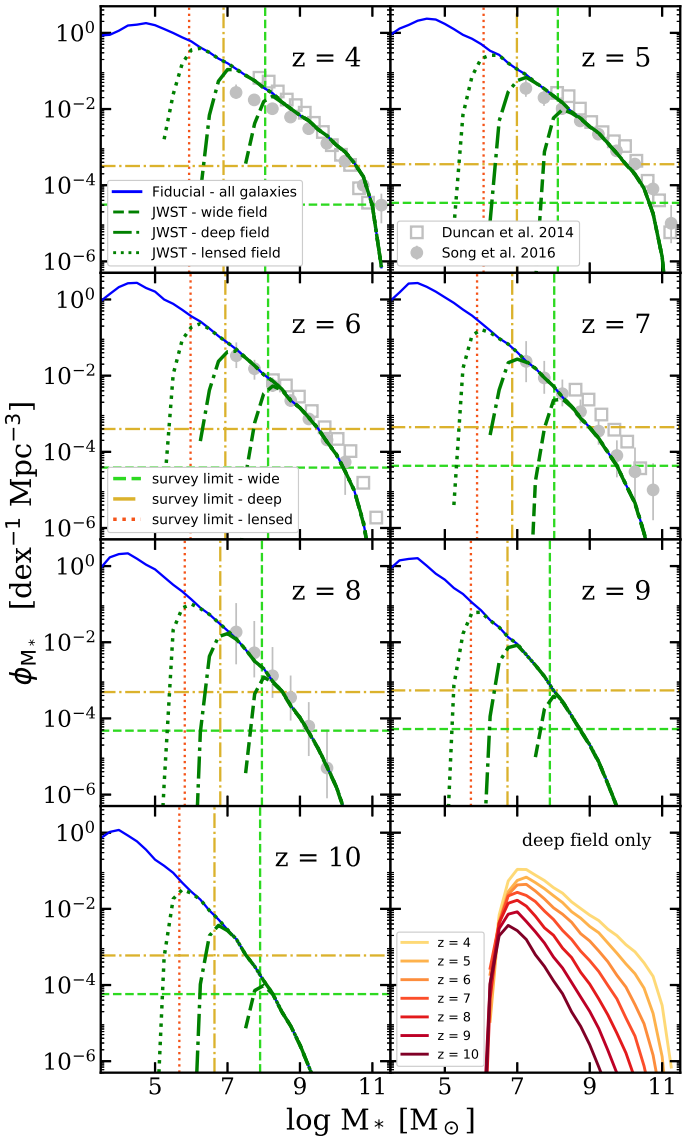

Figure 3.2 Predicted stellar mass functions (SMFs) and their evolution with redshift, based on our fiducial model. Results are shown for all galaxies, and for samples selected to represent wide, deep, and lensed *JWST* surveys. Observational constraints from Duncan et al. (2014) and Song et al. (2016) are shown to guide the eye. The vertical lines mark where the survey completeness reaches 50%, and the horizontal lines mark where we expect one galaxy in the probed volume. The vertical and horizontal lines are colour coded according to the survey area and depth as shown in the legend. See Table 3.1 for assumed survey specifications.

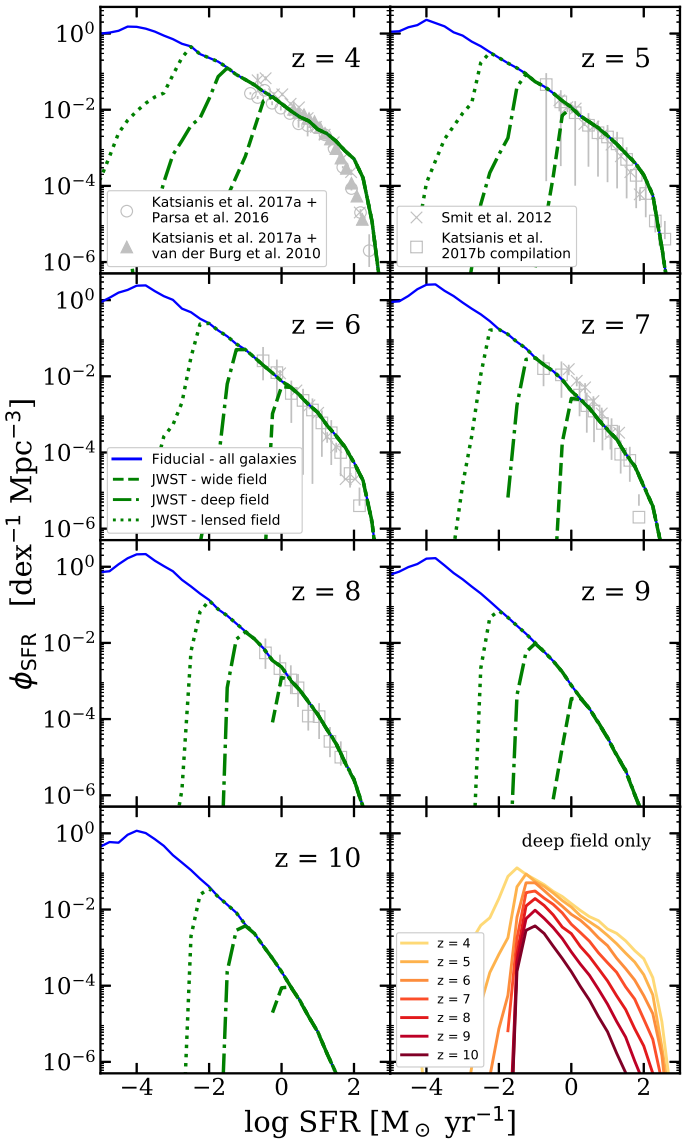

Figure 3.3 Predicted star formation rate functions (SFRFs) and their evolution with redshift. Results are shown for all galaxies, and for samples selected to represent wide, deep, and lensed *JWST* surveys. Observational constraints are shown to guide the eye, as in Figure 3.5. See Table 3.1 for assumed survey specifications.

and other parallel fields. The main differences between their results originate from the assumed normalization and slope of the $M*$–$M_{\rm UV}$ relation, which causes Duncan et al. to consistently find higher stellar masses for galaxies than Song et al. in faint UV bins. These two studies also treat faint objects slightly differently. Duncan et al. fitted their $M*$–$M_{\rm UV}$ relation to a wide stellar mass range down to $\log(M_*/M_\odot) \sim 8$, where stellar masses for galaxies with $\log(M_*/M_\odot) < 9$ are biased toward higher masses, and Song et al. used a hybrid approach where high-mass galaxies are fitted individually and lower-mass galaxies are stacked. Moreover, the Song et al. error bars only include random uncertainties, while the Duncan et al. ones include Poison errors and photometric redshift uncertainties.

Similarly, Fig. 3.5 shows distribution functions for star formation rate (SFRFs) and their evolution across the same redshift range. Our results are compared to observational constraints by Katsianis et al. (2017b) at $z = 4$, Smit et al. (2012) at $z = 4$–7, and Katsianis et al. (2017a) at $z = 5$–8. These constraints on SFR are determined based on a conversion between UV luminosity and SFR (e.g. Kennicutt 1998; Smit et al. 2012). The Katsianis et al. (2017b) study is based on rest-frame UV luminosity functions from Parsa et al. (2016) and van der Burg et al. (2010), both of which are based on a compilation of ground- and space-based deep-field observations. The Katsianis et al. (2017a) results are calculated based on UV LFs presented in Bouwens et al. (2015b). The observational constraints shown have accounted for dust attenuation. Once again, the results from our fiducial model show good agreement with these observational constraints, and are clearly favoured over the model with a constant H_2 depletion time. Note that observational estimates of SFR derived from UV luminosity probe a time-scale of approximately 100 Myr (Kennicutt & Evans 2012). The SFR predictions from the SAM have been averaged over 100 Myr for an appropriate comparison.

As shown in Fig. E.1 in Appendix C, the abundance of halos rises continuously towards lower halo mass. Our simulation routinely samples a wide range of halo masses down to the atomic cooling limit and halo assembly histories are traced down to 100th of the root mass. The 'flattening' or 'turnover' seen at the low-mass / low-SFR end of the distribution functions in our models is a physical prediction resulting from inefficient cooling, rather than from insufficient resolution as in most numerical simulations. However, we note that our model does not include H_2 cooling or metal cooling below 10^4 K. We find that the critical stellar mass where the turnover occurs evolves mildly over redshift, from $\log(\mathrm{M}_*/\mathrm{M}_\odot) \sim 4$ at $z \sim 10$ to $\log(\mathrm{M}_*/\mathrm{M}_\odot) \sim 5$ at $z \sim 4$, while the critical SFR remains nearly constant at $\log(\mathrm{SFR}/(\mathrm{M}_\odot\ \mathrm{yr}^{-1})) \sim -4$ over this period. A cautionary note is that these turnovers are not representing the same galaxies. Instead, each redshift is an independent snapshot that portrays the demographics of the current galaxy population. It is also interesting to

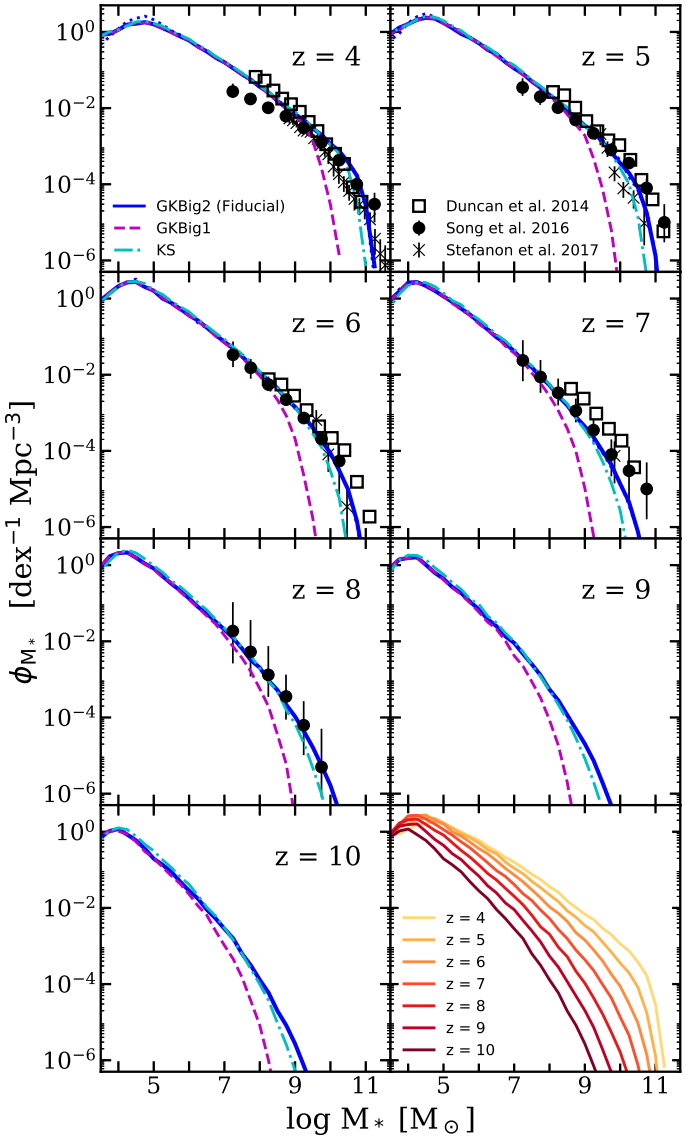

Figure 3.4 Predicted stellar mass functions (SMFs) and their evolution with redshift. The blue solid line shows the results of the GK-Big2 (fiducial) model, the purple dashed line shows the GK-Big1 model, and the cyan dot-dashed line shows the KS model. Our results are compared to observational constraints from Duncan et al. (2014), Song et al. (2016), and Stefanon et al. (2017). We see that the GK-Big2 (fiducial) and KS model are consistent with observations, while the GK-Big1 model does not produce enough massive galaxies. We also show the case where photoionization squelching is turned off for our fiducial model with the blue dotted line. The differences are very subtle and are only visible for the very lowest mass galaxies at $z = 4$ and 5. See text for full explanation.

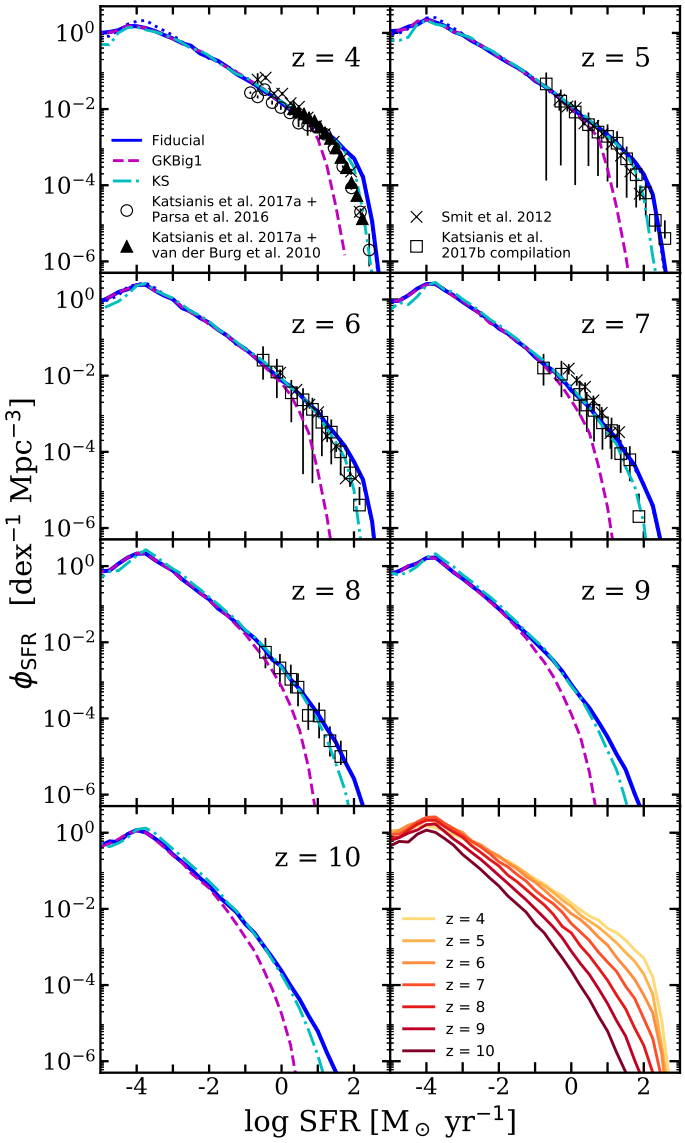

Figure 3.5 Predicted star formation rate functions (SFRFs) and their evolution with redshift. The blue solid line shows the results of the GK-Big2 (fiducial) model, the purple dashed line shows the GK-Big1 model, and the cyan dot-dashed line shows the KS model. Our results are compared to the Katsianis et al. (2017b) study, which is based on observations from Parsa et al. (2016) and van der Burg et al. (2010), and the Katsianis et al. (2017a) compilation, which is calculated based on UV LFs from Bouwens et al. (2015b), and Smit et al. (2012). Once again, the GK-Big2 (fiducial) and KS model are consistent with observations, and the GK-Big1 model does not produce enough rapidly star-forming galaxies. We also show the case where photoionization squelching is turned off for our fiducial model with the blue dotted line. The differences are very subtle and are only visible for the very lowest mass galaxies at $z = 4$ and 5. See text for full explanation.

note that the SFR function is closer to a pure power law at high redshift ($z \gtrsim 6$), and begins to develop an exponential cut-off at high SFR at $z \sim 5$.

For both SMF and SFRF predictions, we have experimented with turning off the feedback from photoionization squelching in our fiducial model, shown as a dotted line in Fig. 3.4 and 3.5. These results show that, at least with the simple implementation adopted in our model, the effect of squelching is negligible across galaxies of all masses, even at $z = 4$. This is because the characteristic mass from the Okamoto et al. (2008) simulations is getting very close to the atomic cooling limit. Thus only the very low mass halos are affected by squelching. Additionally, we have tested the effect of switching off AGN feedback in our models, and find that it has no discernible effect on the predictions in this redshift range (see the discussion in Chapter 2).

In a similar spirit as the experiments done in Chapter 2, we systematically varied the stellar feedback slope $\alpha_{\rm rh}$ and the SF time-scale $\tau_{*,0}$ within a sensible range in an attempt to quantify the effects of these uncertainties in the key model parameters. Here, we show low-mass and massive galaxy populations separately in Fig. 3.6 and Fig. 3.7, respectively.

We showed in Chapter 2 that the low rest-frame UV luminosity galaxy populations are very sensitive to the efficiency of stellar driven winds. Here, we show a similar effect for stellar mass, where we explore four values of $\alpha_{\rm rh} = 2.0$, 2.4, 3.2, and 3.6, where larger values of $\alpha_{\rm rh}$ imply a steeper dependence of mass loading factor on galaxy circular velocity, resulting in more gas being ejected from galaxies in low mass halos (see §2.3 in Chapter 2 for details). Note that these alternative values will not necessarily reproduce observations at $z \sim 0$; we are adopting these values as an attempt to explore alternative scenarios where the strength of feedback may have been stronger or weaker at early times. In our model, feedback from stellar winds is modelled using a physically motivated, redshift-independent recipe. This feedback mechanism could be sensitive to many intertwined physical properties of galaxies, which could lead indirectly to an effective evolution with cosmic time. The feedback efficiency is poorly constrained especially in low-mass galaxies and at high redshifts due to the challenges of obtaining direct observational constraints. Henceforth, without precluding the possibility of having an effectively evolving $\alpha_{\rm rh}$, we have experimented with adopting different constant values of $\alpha_{\rm rh} = 2.8 \pm 0.4$ and ± 0.8. On the other hand, $\tau_{*,0}$ effectively acts as a normalizing factor for the SF relation, which converts the surface density of molecular hydrogen $\Sigma_{\rm H_2}$ to the surface density of SFR $\Sigma_{\rm SFR}$ (see §2.1 in Chapter 2 for details). Larger $\tau_{*,0}$ represents a longer SF time-scale or, effectively, a longer gas depletion time-scale, yielding a lower star formation rate per unit surface H_2 density, and vice versa. To explore the effect of uncertainties in the observationally-motivated, empirical SF recipe, we experiment with increasing and decreasing $\tau_{*,0}$ by a factor of two from its fiducial value of unity (which corresponds to the observed normalization).

In Fig. 3.6, we see that the low-mass-end slope $\alpha_{\rm rh}$ becomes steeper when $\alpha_{\rm rh}$ is decreased to 2.4 and 2.0, due to the less efficient mass ejection by stellar feedback. The slope flattens when $\alpha_{\rm rh}$ is increased to 3.2 and 3.6. In the least massive populations, we see a shift in the 'turnover' mass when we vary the feedback efficiency. This is because the turnover in our models always occurs roughly at a fixed halo circular velocity (temperature) corresponding to the atomic cooling limit, but changing the feedback prescription changes the relationship between halo mass or velocity and galaxy stellar mass. We also find that the low-mass populations are insensitive to the adopted value of $\tau_{*,0}$ (see SPT15 for a detailed discussion of the physical reasons for this).

Similarly, in Fig. 3.7, we explore the impact of adopting a range of values for $\alpha_{\rm rh}$ and $\tau_{*,0}$ on the massive galaxy populations. In this regime, the results are more degenerate, showing significant dependence on both parameters, although SF time-scale has a stronger effect on more massive galaxies and the SN feedback has a greater effect on low-mass objects. We note that tuning $\tau_{*,0}$ alone to alter the bright end prediction will result in a change in predicted gas fraction. Thus observational constraints on gas content in galaxies at high redshift can help to break these degeneracies.

As noted before, the current observational constraints on stellar masses of faint objects at high redshift are highly uncertain, as reflected in the discrepancies in current estimates in the literature. Our models can easily accommodate the range of scenarios presented by current observations with variations in parameters that are well within the observational and theoretical uncertainties. Our results illustrate how future measurements with *JWST* and wide-field surveys with LSST and WFIRST will be complementary in constraining different physical processes in galaxy formation at extreme redshifts.

3.3.3 Comparison with other models

In this subsection, we compare the SMFs and SFRFs predicted by our fiducial model to a representative collection of theoretical predictions, including empirical models, semi-analytic models, and cosmological hydrodynamic simulations from the literature. First, we very briefly summarize the specifications of these models and simulations. We note that these models adopt different cosmological parameters, very different approaches for modelling the baryonic physics, and different approaches for calibration. We compare the results at face value without attempting to correct for any of these differences. Furthermore, providing the full details about these models or attempting to understand the sources of differences in their predictions is beyond the scope of this work. We refer the reader to the included references for each of these simulations for full details.

A number of numerical simulations are included in our comparison; BlueTides is a large-volume cosmological hydrodynamic simulation that focuses on the high-redshift universe, with a box that

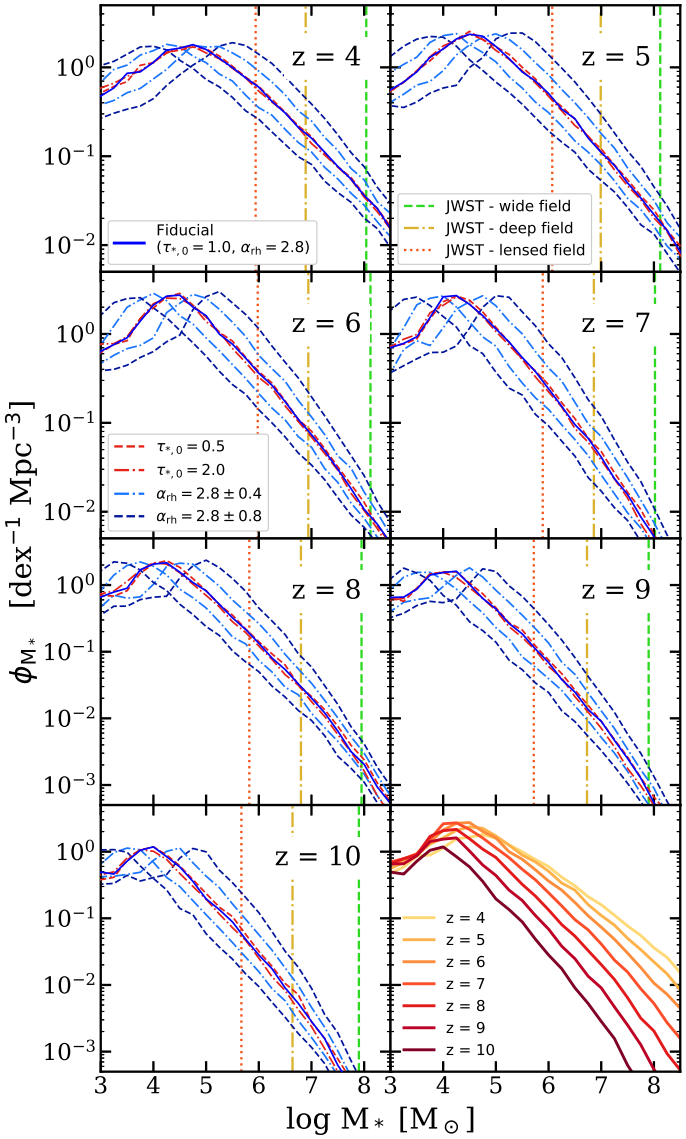

Figure 3.6 The low-mass end of the predicted stellar mass functions (SMFs) and their evolution with redshift. The blue solid line shows the results from our fiducial model. Red lines represent the cases where we alter the SF time-scale $\tau_{*,0}$, where dashed and dot-dashed lines are $\tau_{*,0} = 0.5$ and $\tau_{*,0} = 2.0$, respectively. The light-blue lines show cases where the mass-loading factor of stellar driven winds $\alpha_{\rm rh}$ is altered, where $\alpha_{\rm rh} = 2.8 \pm 0.4$ (dot-dashed) and ± 0.8 (dashed). Blue dot-dashed lines show the cases where we let $\alpha_{\rm rh} = 2.4$ (above) and 3.2 (below), and light blue dashed lines show the cases where we let $\alpha_{\rm rh} = 2.0$ (above) and 3.6 (below). Red lines represent the cases where we increase or decrease $\tau_{*,0}$ by a factor of 2 from its fiducial value of unity; dashed and dot-dashed lines are $\tau_{*,0} = 0.5$ and $\tau_{*,0} = 2.0$, respectively. The vertical dashed lines represent survey completeness of 50% for example *JWST* surveys similar to legacy *HST* counterparts; see Table 3.1 for details. The last panel summarizes the evolution of the low-mass end of the SMFs predicted by the fiducial model. We see that both the slope of the low-mass SMF and the location of the turnover are strongly affected by the model for stellar driven winds, but are not significantly affected by variations in the SF time-scale.

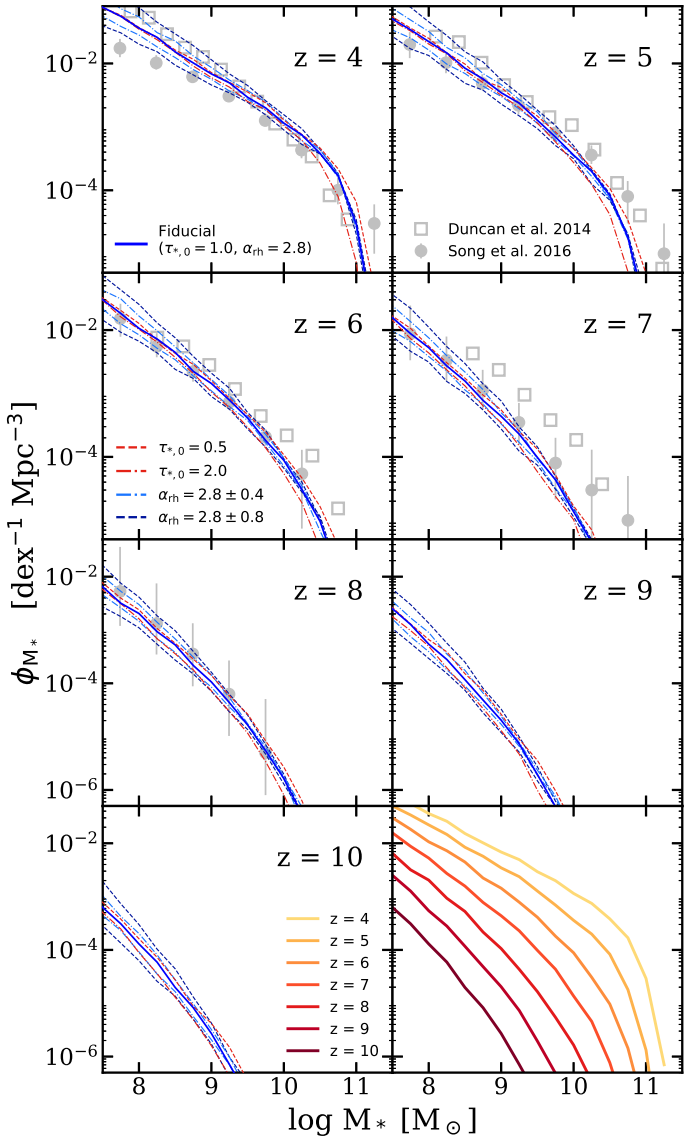

Figure 3.7 The massive end of the predicted stellar mass functions (SMFs) and their evolution with redshift. See caption of Fig. 3.6 for legend details. The gray circle and open square symbols show observational constraints from Duncan et al. (2014) and Song et al. (2016), respectively. The last panel summarizes the evolution of the massive end of the SMFs predicted by the fiducial model. Varying the SF time-scale has the greatest impact on very massive galaxies at high redshift, while varying the stellar wind model has a smaller effect in this regime.

is 400 Mpc h^{-1} on a side, resolving galaxies with $M_* \gtrsim 10^8 \mathrm{M}_\odot$ towards the end of their simulation, which stops at $z \sim 8$. They presented predictions for both SMFs and SFRFs (Wilkins et al. 2017). The Illustris simulation has a box 106.5 Mpc on a side with dark matter particle of mass $6.3 \times 10^6 \mathrm{M}_\odot$ (Genel et al. 2014). The Evolution and Assembly of Galaxies and ther Environment (EAGLE) simulations is a suite of cosmological hydrodynamical simulations of simulated volumes ranging from 25 to 100 cMpc at various mass resolution (Schaye et al. 2015). In our comparison, we used the stellar mass functions from their Ref-L100N1504 and Recal-L025N0752 runs, which have box sizes of 100 cMpc and 25 cMpc on a side and dark matter particle mass of 1.21×10^6 $\mathrm{M}_\odot$ and 9.70×10^6 $\mathrm{M}_\odot$, respectively (Furlong et al. 2015). The Feedback in Realistic Environments (FIRE) simulations are a suite of 'zoom-in' simulations extracted from cosmological volume simulations, for which regions are 'resimulated' at higher resolution with detailed physical processes incorporated (Hopkins et al. 2014). We use SMFs from the FIRE-2 simulations (Hopkins et al. 2018; Ma et al. 2018b), which are obtained by weighting the results from the zoom-ins with cosmological halo mass functions. Due to space limitations, other simulations, such as the Renaissance Simulations (O'Shea et al. 2015), the FirstLight project (Ceverino, Glover, & Klessen 2017; Ceverino, Klessen, & Glover 2018, 2019), and SPHYNX (Cabezón, García-Senz, & Figueira 2017), as well as semi-analytic model results from Dayal et al. (2014), are omitted from our comparison.

We also compare with predictions from the Dark-ages Reionization and Galaxy formation Observables from Numerical Simulations (DRAGONS) project, which consists of a SAM MERAXES (Mutch et al. 2016) that is built on top of the *Tiamat* suite of N-body simulations of dark matter particle of mass $\sim 2.64 \times 10^6$ h^{-1} Mpc (Poole et al. 2016). We show the SMFs presented in Qin et al. (2017).

We include comparisons with several sub-halo abundance matching (SHAM), empirical, and semi-empirical models. The UNIVERSEMACHINE (Behroozi et al. 2019) obtained dark matter halo populations and properties from the Bolshoi-Planck dark matter simulations (Klypin et al. 2016; Rodríguez-Puebla et al. 2016), which simulate a volume of 250 Mpc h^{-1} on a side containing 2048^3 dark matter particles of $1.5 \times 10^8 \mathrm{M}_\odot$. Within these dark matter halos, galaxy populations are constructed to fit the SFR–v_{Mpeak} relation, quenching–v_{Mpeak} relation, quenching/assembly history correlation in the local universe ($z \sim 0$), and dust content at high redshifts ($z \sim 4$–10) from a wide range of observational constraints. In this work we use the UNIVERSEMACHINE Early Data Release (EDR) catalogues[1]. The Tacchella et al. (2018) model uses dark halo merger trees extracted from the *Copernicus complexio Low Resolution* (COLOUR) simulations (Hellwing et al. 2016; Sawala et al. 2016), which contain 1620^3 dark matter particles of 6.196×10^6 $\mathrm{M}_\odot h^{-1}$ in a periodic volume of

[1] http://behroozi.users.hpc.arizona.edu/UniverseMachine/EDR/umachine-edr.tar.gz

70.4 Mpc h^{-1} on a side. The only observational constraint this model is calibrated to is the UV LF observed at $z = 4$, where the star-formation efficiency is treated as a free parameter. Their results are then checked against the measured cosmic star formation rate, $M_{\rm UV}$–M_* relation, stellar mass function, etc. The JADES Extragalactic Ultra-deep Artificial Realization (JAGUAR; Williams et al. 2018) model provides mock catalogues specifically for upcoming *JWST* observations. JAGUAR makes no assumptions about the underlying halo populations. Instead, it directly maps observable properties, such as $M_{\rm UV}$, to a number of physical properties using simple relations from observations, such as the $M_{\rm UV}$–M_* relation, $\beta_{\rm UV}$–$M_{\rm UV}$ relation, and spectroscopic properties using SED fitting to *3D-HST* catalogues (Skelton et al. 2014; Momcheva et al. 2016). The semi-empirical model Emerge (Moster et al. 2018) is very similar to the UniverseMachine models but is not included in our comparison, again for reasons of space limitations.

In Fig. 3.8, we compare our predicted SMFs to the predictions of Illustris, Eagle, DRAGONS, BlueTides, UniverseMachine, Tacchella et al., and Williams et al.. We note that some of these simulated SMFs are based on the Salpeter (1955) IMF and the Kroupa (2001) IMF, and are converted to Chabrier IMF by adding -0.21 and -0.03 dex, respectively (Madau & Dickinson 2014). In general, most of these predictions are within 0.5 dex from our predictions within the mass range where objects are well resolved (in the low-mass end) and well sampled (in the massive end) in the given simulation. We also note that the excellent agreement for $M_* \gtrsim 9$ at $z \sim 4$–7 may be to some extent by construction, since all simulations are matching the available observational constraints either actively (by calibration) or passively (used as crosschecks). Our models agree quite well overall with the predictions from numerical hydrodynamic simulations, including BlueTides, Illustris, and FIRE, over a broad stellar mass and redshift range. It is intriguing that these models, which produce converging predictions on the massive end, are producing rather different predictions for the low-mass populations. For instance, the most optimistic model, DRAGONS, and the least optimistic model, Williams et al. predict the abundance of $M_* = 10^7 {\rm M}_\odot$ galaxies differs by > 1 dex at $z = 7$, and the difference becomes even larger at low redshifts. The comparison also shows that there is a shortfall of massive objects in the JAGUAR predictions. Note that in this redshift range, the Williams et al. model does not explicitly match their SMFs to observations, but rather forward modelled them to match UV LFs adopting simple assumptions for the M_*–$M_{\rm UV}$ and $\beta_{\rm UV}$–$M_{\rm UV}$ relations based on local observations.

The SMFs at high redshift predicted by the UniverseMachine tend to have a higher normalization than previous results due to improved treatment of star formation history priors, which require more stellar mass especially at $z \sim 8$ (P. Behroozi, private communication). On the other hand, the lower mass limit in halo mass in the UniverseMachine is $\sim 10^{10.5}$–$10^{10.7}{\rm M}_\odot$ below

which merger trees are not sufficiently resolved to result in accurate galaxy properties. We note that both the Tacchella et al. model and the UniverseMachine are not explicitly calibrated to match observed SMF constraints at $z > 4$. Due to the specific set of assumptions and choices made in their models, though both of these models are able to reproduce observations in the low-redshift universe, their predictions in the poorly constrained high-redshift differ significantly. Investigating and understanding these differences in detail is beyond the scope of this work, and we refer the reader to the relevant papers for full details.

Similarly, in Fig. 3.9, we compare our predicted SFRFs to Illustris, BlueTides, UniverseMachine, and Tacchella et al.. We note that some of the SFRs predicted by these simulations are based on the Salpeter (1955) IMF and the Kroupa (2001) IMF, and are converted to Chabrier IMF by adding -0.20 and -0.03 dex, respectively, for this comparison (Madau & Dickinson 2014). At first glance, predictions by Illustris and Tacchella et al. seem to agree with predictions from our fiducial model quite well for the abundance of galaxies with SFR ~ 0.1 $\mathrm{M_\odot\ yr^{-1}}$ across redshifts. However, they diverge for SFR $\gtrsim 1$ $\mathrm{M_\odot\ yr^{-1}}$. The SFR predicted by UniverseMachine at $z \lesssim 7$ are consistently lower than other predictions by ~ 0.5 dex and are higher at $z \gtrsim 9$ for similar reasons to those noted above.

In summary, it is encouraging that physically motivated, *a priori* models make similar predictions for these fundamental quantities at redshifts where they were not calibrated to observations, in spite of containing very different modelling approaches to baryonic processes. The differences between these models and empirical and semi-empirical models may yield interesting insights into the relationship between galaxy observable properties and halo properties. We pursue this further in a later section of this chapter.

3.3.4 Cold gas content of high-redshift galaxies

We learned that both cold gas and molecular gas are important tracers for star formation activity from observing nearby galaxies (e.g. Kennicutt 1998; Bigiel et al. 2008). Empirical relations based on these observations are then extensively incorporated in galaxy formation models, including both conventional numerical simulations and semi-analytic models. Therefore, in addition to M_* and SFR, the distributions of cold gas mass ($M_{\rm cold}$) and molecular gas mass ($M_{\rm H_2}$) and their evolution can also provide important constraints for disambiguating stellar feedback processes. Historically, it has been somewhat challenging for models to simultaneously match the observed gas fraction, stellar fraction, and stellar metallicity (see SD15 for a detailed discussion).

The gas depletion time-scale is the time-scale for converting all the mass in the cold gas reservoir into stars. The three SF recipes adopted in our SAM can be effectively understood as different

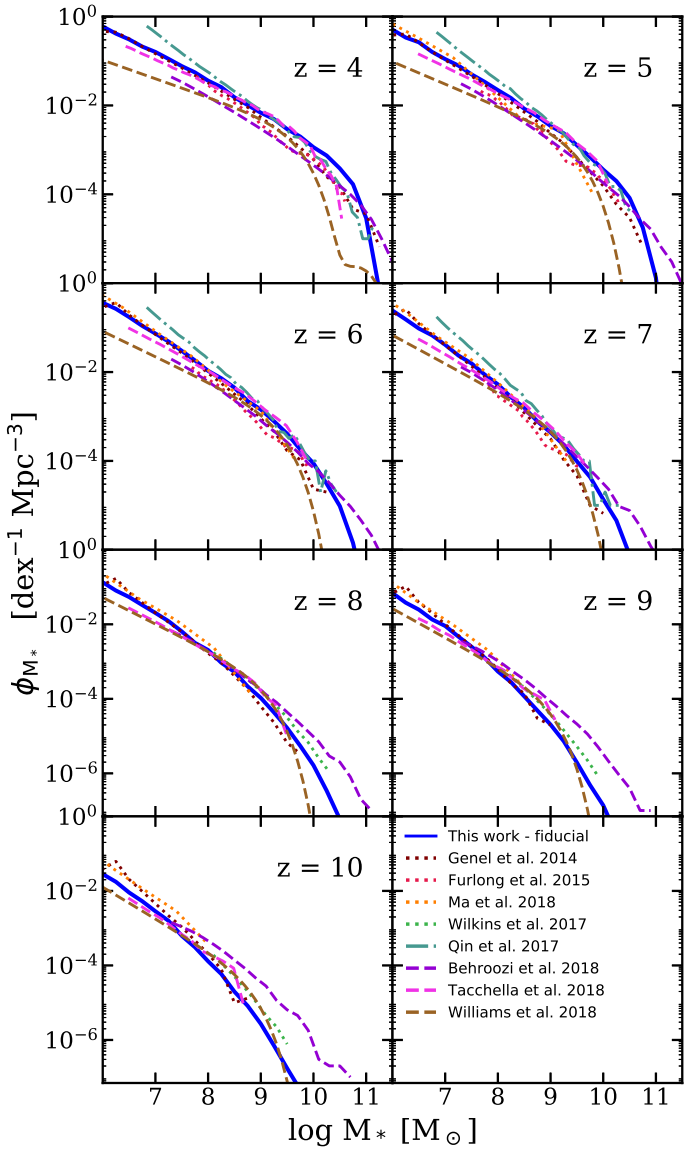

Figure 3.8 Redshift evolution of the SMF between $z = 4$–10 predicted by our fiducial model, compared to a compilation of other theoretical studies, including empirical models UniverseMachine (Behroozi et al. 2019), Tacchella et al. (2018), and JAGUAR (Williams et al. 2018); numerical simulations from Illustris (Genel et al. 2014), Eagle (Furlong et al. 2015), and BlueTides (Wilkins et al. 2017), and semi-analytic models from DRAGONS (Qin et al. 2017). See text for details. Overall, the agreement between different methods is fairly good.

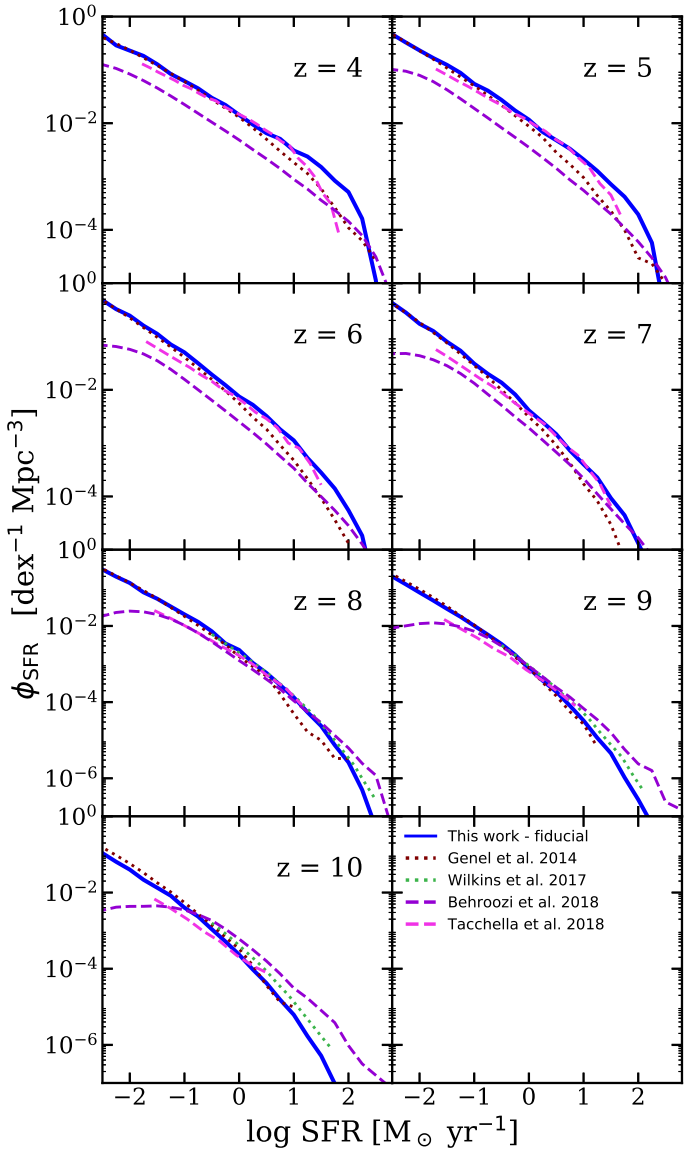

Figure 3.9 Redshift evolution of the SFRF between $z = 4 - 10$ predicted by our fiducial model, compared to a compilation of other theoretical studies, including empirical models UniverseMachine (Behroozi et al. 2019) and Tacchella et al. (2018), numerical simulations from Illustris (Genel et al. 2014), Eagle (Furlong et al. 2015), BlueTides (Wilkins et al. 2017). See text for details.

redshift evolution scenarios for gas depletion time. For instance, the GK-Big1 model corresponds to a constant H_2 depletion time and the GK-Big2 model predicts a gas depletion time that changes with time as the galactic discs gradually become less dense over time. The evolution of these quantities predicted by our models at $z \sim 0$–6 have been studied in depth in SPT15 and PST14, which found that these models produce results that are in qualitative agreement with observed evolution in galaxies at $z \lesssim 6$. In this work, we extend the investigation to $z = 10$. Though constraints on the gas content of galaxies at such extreme redshifts are not currently available, future observational programs and observational facilities may be able to obtain them. Moreover, examining the changes in the predictions for gas content provides complementary insights into our model results.

In this subsection, we show and compare the one-point distribution functions of cold gas and molecular gas masses among our three star formation models. Throughout this work, cold gas mass is defined as $M_{\rm cold} = 1.4(M_{\rm HI} + M_{\rm HII} + M_{\rm H_2})$, where the factor of 1.4 is the correction for the mass of He. In Fig. 3.10, we compare the cold gas mass functions (CGMF) for galaxies across the three SF models. Note that these models are calibrated to match the SMF, gas fraction, and other observables at $z \sim 0$. Overall, there are relatively more low-$M_{\rm cold}$ galaxies in the H_2-based SF models (GK-Big1 and Fiducial) compared to the $M_{\rm cold}$-based KS model due to the more efficient scaling between cold gas to SF in low-mass halos (see Fig. 1 in SPT15). Given that the low-M_* and low-SFR galaxy populations are nearly identical as shown in Fig. 3.4 and 3.5, we can infer that gas depletion has been much slower in low-mass galaxies in H_2-based models. The difference narrows towards low redshift. This is largely due to the adopted metallicity dependent formation efficiency of H_2 in the GK-based models — H_2 formation is less efficient in low-mass galaxies at high redshifts, because of the lower metallicities in both cases.

In Fig. 3.11, we show the H_2 mass function (H_2MF) and take a closer look at the difference between the Fiducial and the GK-Big1 models. Note that predictions for molecular gas content are only available in models with multiphase gas partitioning (the ones labelled GK). Recall that these two models are very similar, except that the SF relation slope in the fiducial model steepens to ~ 2 at higher surface density, as suggested by some observational evidence, while it remains constant in the GK-Big1 model. Thus, the differences in H_2 mass between these models allow us to indirectly probe the dependence of star formation on H_2 density. We see that the fiducial model has fewer galaxies with high H_2 content at all redshifts studied, but the difference appears relatively subtle.

We further investigate the differences in these models by looking into the cold gas fraction, which is defined as $f_{\rm gas} \equiv M_{\rm cold}/M_*$, shown in Fig. 3.12. For each model, we show the scaling relation between $f_{\rm gas}$ and M_* for all three of our SF models, where the solid lines mark the median and the dashed lines mark the 16th and 84th percentile. As mentioned before, the H_2-based models have

significantly higher gas fractions especially in low-M_* galaxies, because of the low H_2 content of these galaxies and resulting low SF efficiency. Similarly, we also show the molecular gas fraction, defined as $f_{H_2} \equiv M_{H_2}/M_*$, shown in Fig. 3.13. This figure perhaps most clearly illustrates how the fiducial and GK-Big1 models differ.

As discussed previously, and shown in SPT15, high M_* galaxies have longer H_2 depletion times (lower efficiencies for converting molecular gas into stars) in the GK-Big1 model than in the fiducial model at high redshift, and the time-scale for converting H_2 into stars is the rate limiting factor for forming stars in these galaxies. Therefore, we see much larger reservoirs of 'leftover' molecular gas in the GK-Big1 model, also explaining why it fails to produce as many galaxies with large stellar masses at high redshift.

To briefly summarize this entire section, we find that SMFs and SFRFs predicted by our fiducial model agree well with available observations, within the large observational uncertainties. In both cases, the choice of star formation recipe mainly affects the most massive and rapidly star-forming galaxies. Varying the efficiency of stellar feedback shifts the location of the low-mass turnover and changes the slope of the low-mass end of the SMF. Varying the star formation time-scale (normalization of the SF recipe, or gas depletion time) mainly impacts the massive end of the SMF at very high redshifts, where star formation has not yet become self-regulated. The cold gas and H_2 content of high redshift galaxies is quite sensitive to the details of the star formation recipe, and can help break degeneracies between star formation efficiency and stellar driven wind parameters. For example, the cold gas fraction can discriminate between a model with high star formation efficiency and strong ejective feedback, vs. a model with low star formation efficiency and weaker feedback.

3.4 Scaling Relations for High-Redshift Galaxy Populations

In this section, we investigate the scaling relations among galaxy properties for the high-redshift populations predicted by our fiducial model that are expected to be detected by *JWST*.

In Fig. 3.14 to 3.20, we present the distributions between every permutation among a few selected photometric and physical properties at $z = 6$, 8, and 10, respectively, for galaxies that are detectable given the very optimistic detection limit for lensed surveys ($m_{\rm F200W} \lesssim 34.0$). These plots are prepared using the `corner.py` module provided by Foreman-Mackey (2016). Plots for other redshifts are omitted to avoid clutter and are made available online. The properties included here are halo mass ($M_{\rm H}$), scale radius of the stellar disc (R_*), cold gas metallicity ($Z_{\rm cold}$), SFR, M_*, dust attenuated rest-frame UV luminosity ($M_{\rm UV}$), and observed-frame IR magnitude in the NIRCam F200W filter $m_{\rm F200W}$. The diagonal panels show one-dimensional histograms for the property labelled on

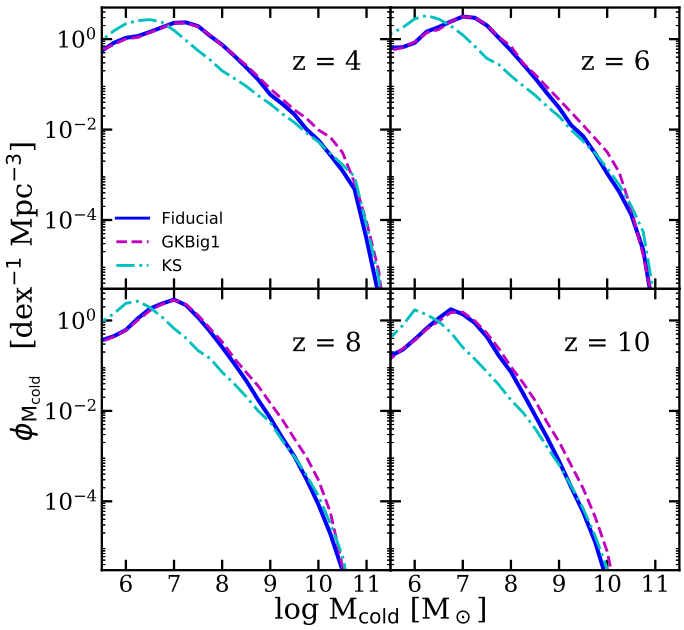

Figure 3.10 Predicted distribution functions for cold gas $M_{\rm cold}$ at $z = 4$, 6, 8, and 10. The blue solid line shows the results of the GK-Big2 (fiducial) model, the purple dashed line shows the GK-Big1 model, and the cyan dot-dashed line shows the KS model. While these models all made very similar predictions for the low-mass end of the SMF, they make rather different predictions for the cold gas mass function at low gas masses.

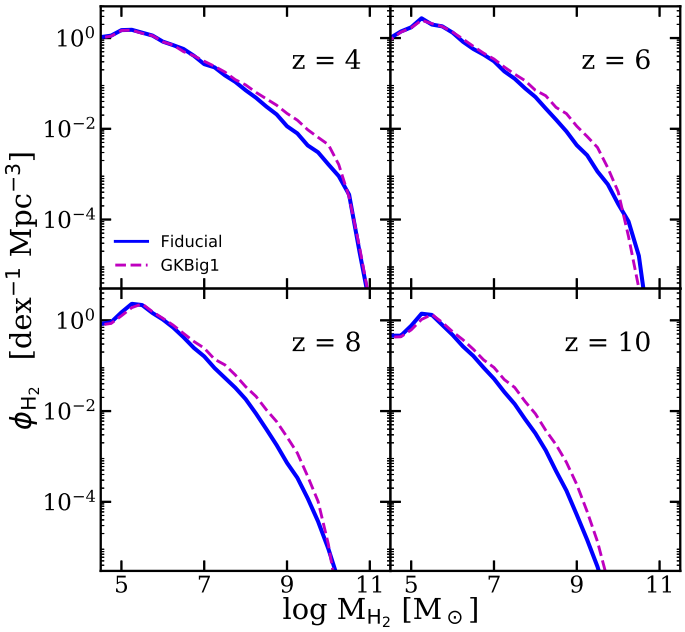

Figure 3.11 Predicted distribution functions for molecular hydrogen $M_{\rm H2}$ at $z = 4$, 6, 8, and 10. The blue solid line shows the results of the GK-Big2 (fiducial) model and the purple dashed line shows the GK-Big1 model. These models yielded very different predictions for the SMF, but predict similar $M_{\rm H2}$ mass functions.

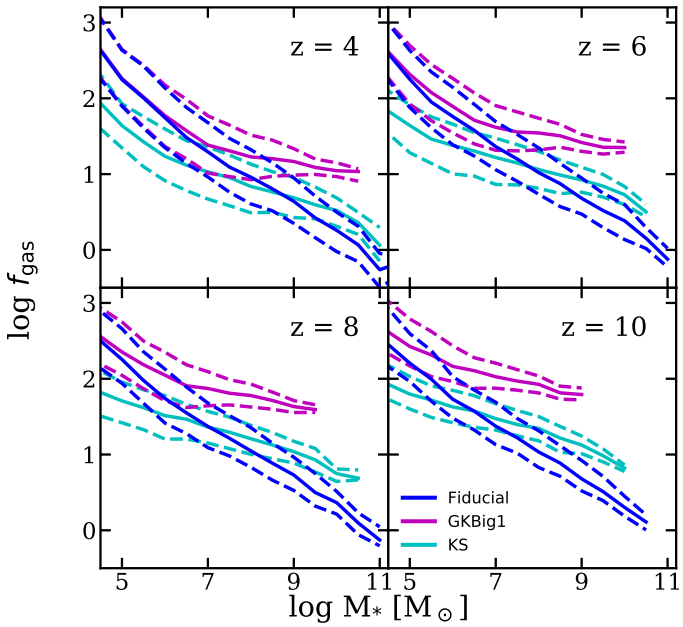

Figure 3.12 Predicted relationship between stellar mass and cold gas fraction $f_{\rm gas} = M_{\rm cold}/M_*$ at $z = 4$, 6, 8, and 10. The solid line marks the median and the dashed lines mark the 16th and 84th percentile. The blue lines show the results of GK-Big2 (fiducial) model, the purple dashed lines show the GK-Big1 model, and the cyan dot-dashed lines show the KS model. Varying the star formation recipe yields dramatically different predictions for the cold gas content of galaxies at high redshift.

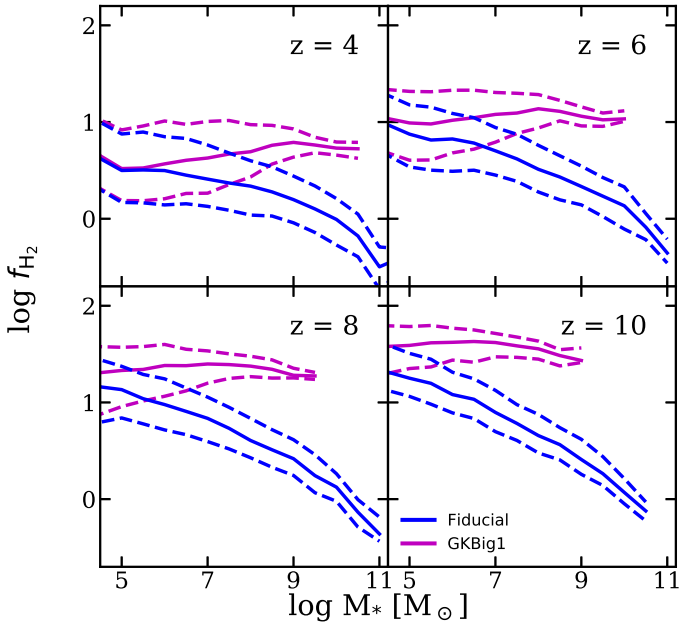

Figure 3.13 Predicted relationship between stellar mass and molecular gas fraction $f_{\rm H2} = M_{\rm H2}/M_*$ at $z = 4$, 6, 8, and 10. The solid line marks the median and the dashed lines mark the 16th and 84th percentiles. The blue lines show the results of the GK-Big2 (fiducial) model and the purple dashed lines show the GK-Big1 model. The $M_{\rm H2}$ content of galaxies is very sensitive to the star formation recipe.

the corresponding x-axis, with the 16th, 50th, and 84th percentiles of the distribution marked by the vertical lines. Each of the off-diagonal panels shows the distribution between two properties using a two-dimensional hybrid histogram-scatter plot, where the 2D histograms are weighted by the halo abundances. Objects with abundances that fall below the 16th percentile are plotted as individual points. We also mark the detection limits assumed for *JWST* wide surveys ($m_{\rm F200W} \sim 28.6$) and deep surveys ($m_{\rm F200W} \sim 31.5$) in these plots. Note that the turnovers in the histogram for halo mass and other quantities are not from the resolution limit of our model, but rather due to the intrinsic scatter in galaxy and halo properties at the detection limit.

These diagrams can be used in several ways. One way is for observers to be able to quickly estimate the range of expected physical properties for galaxies with a given observed-frame magnitude. To use the plots in this way, one can look across the columns at a given value of $m_{\rm F200W}$. Alternatively, simulators or modellers can use them to estimate the expected physical and observable properties of galaxies hosted by halos of a given mass. To use the plot in this way, one can look up the rows at a given value of $M_{\rm H}$. In addition to the normalization and slope of the scaling relations between various quantities, the expected scatter in different quantities is also very important. We notice that some relations (such as stellar mass vs. SFR) are very tight, while others (such as all relations with disc size), show almost no correlation. Some relations are linear and monotonic, while others have a break (such as halo mass vs. stellar mass, SFR, and rest-UV luminosity). These scaling relations can also be used to guide and aid in the comparison with semi-empirical models.

We see fairly tight correlations between $m_{\rm F200W}$ and $M_{\rm UV}$, SFR, and M_*. We track the median of these relations and their redshift evolution, and find that they remain linear and evolve steadily between $z = 4$–10. Fitting functions for selected scaling relations are presented in Table D.1 in Appendix D. Note that the correlations are disrupted in the bright, massive populations due to the effect of dust attenuation, which is not accounted for in the fitting. The median values of $M_{\rm UV}$, SFR, and M_* specifically corresponding to the wide- and deep-field detection limits across redshifts are then presented in Table 3.2. These corresponding values are also marked in Fig. 3.16, 3.18, and 3.20 with lines of matching colour and style.

As illustrated in the last panel of Fig. 3.2 and 3.3, it is interesting to note that the limiting M_* and SFR for a given observed-frame detection limit seems to be evolving only slightly across $z = 4$–10. Intuitively, one would expect a more rapid evolution towards higher M_* and SFR at higher redshift, because of the dimming due to the larger luminosity distances. However, our results demonstrate otherwise. As shown in Table 3.2, for galaxies with a certain observed-frame IR magnitude, high-redshift galaxies seem to have higher rest-frame UV luminosities, lower stellar masses, and lower SFRs compared to their low-redshift counterparts. Equivalently, we can say that

galaxies at high redshift are intrinsically brighter than their low-redshift counterparts of similar rest-frame UV luminosities, stellar masses, or SFRs (see also Fig. 3.24). This effect is apparently nearly sufficient to cancel out the dimming effect.

Table 3.2 $M_{\rm UV}$, M_*, and SFR for galaxies at the detection limit of a representative *JWST* wide survey ($m_{\rm F200W,lim} = 28.6$) and deep survey ($m_{\rm F200W,lim} = 31.5$).

	$M_{\rm UV}$	$\log M_*$ [M$_\odot$]	log SFR [M$_\odot$ yr^{-1}]
z	Wide | Deep	Wide | Deep	Wide | Deep
4	-17.03 | -14.40	7.92 | 6.75	-0.73 | -1.82
5	-17.76 | -15.10	7.99 | 6.83	-0.42 | -1.53
6	-17.97 | -15.32	7.92 | 6.76	-0.32 | -1.43
7	-18.11 | -15.47	7.84 | 6.68	-0.27 | -1.38
8	-18.23 | -15.56	7.78 | 6.61	-0.21 | -1.35
9	-18.34 | -15.65	7.69 | 6.53	-0.22 | -1.36
10	-18.32 | -15.68	7.57 | 6.44	-0.27 | -1.39

3.5 Redshift Evolution of Scaling Relations

In this section, we focus on the redshift evolution of scaling relations and other physical properties across $z = 4$–10. In Fig. 3.21, we show the stellar-to-halo mass relation (SHMR) predicted by our fiducial model and compare it to predictions from other models, including semi-empirical model predictions from Tacchella et al. and UniverseMachine. The Tacchella et al. model assumed a Salpeter IMF, and as above we have applied a correction to the stellar masses throughout this section to make them more appropriately comparable to a Chabrier IMF, used in our work. For our predictions, we show the median and the intrinsic scatter of the distributions (16th and 84th percentiles). For the Tacchella et al. and UniverseMachine model outputs, the statistical errors are shown. In the last panel we show the median of the relations predicted at $z = 4$, 7, and 10 to highlight the evolution.

Note that Tacchella et al. assumed a WMAP-7 cosmology, while UniverseMachine assumes Planck 2015 cosmology that is consistent with the one adopted in this work. As shown in Fig. E.1 in Appendix C, changing from the WMAP-7 to Planck15 cosmology increases the abundance of halos across all masses, and thus effectively shifts the halo occupation function. Thus halos of the same mass must contain galaxies of lower stellar mass in the Planck15 than in WMAP-7 cosmology. It is intriguing that the results from these models are very consistent at $z = 8$ and 9, but then diverge substantially at lower redshift. We speculate that this is due to both the differences in assumed cosmology and in the observational constraints used by these models. As shown in the lower panel of Fig. E.1, as we move from the legacy WMAP cosmology to the more recent Planck cosmology,

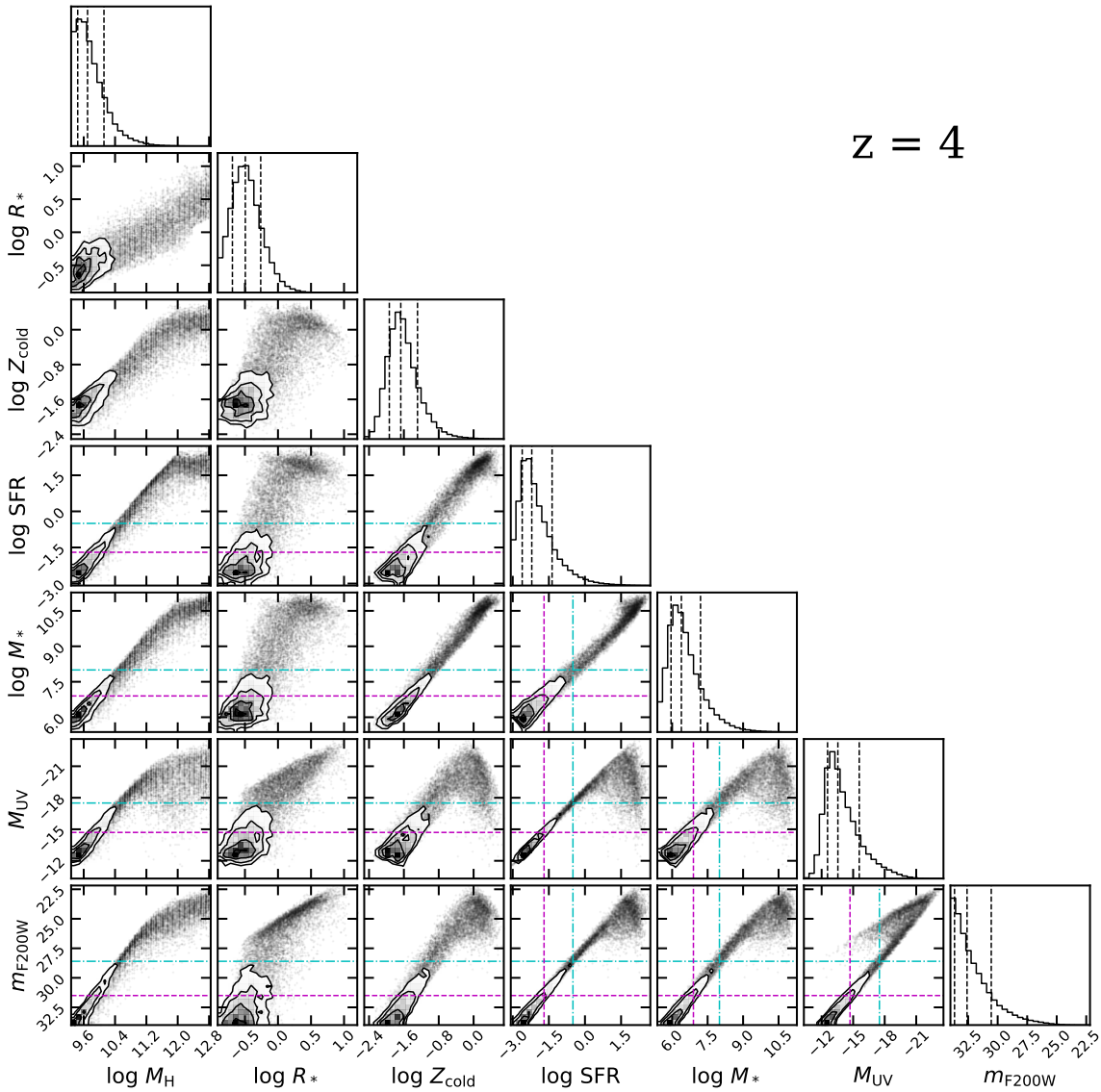

Figure 3.14 We show relations between halo mass $M_{\rm H}$, stellar disc scale radius R_*, cold gas phase metallicity $Z_{\rm cold}$, SFR, stellar mass M_*, rest-frame $M_{\rm UV}$ (with dust attenuation), and $m_{\rm F200W}$ at $z = 4$ for galaxies above the detection limit of *JWST* lensed surveys ($m_{\rm F200W} \lesssim 34.0$) predicted by the fiducial model. All masses are in solar units and R_* is in units of physical kpc. The diagonal panels show histograms for the quantities marked on the corresponding x-axis, where the vertical dashed lines mark the 16th, 50th, and 84th percentiles. The off-diagonal panels show the distribution between two properties using a two-dimensional hybrid histogram-scatter plot, which is colour-coded for the relative object abundances among the bins. The contours mark the 16th, 50th, and 84th percentiles, while objects falling below the 16th percentile are plotted as individual points. The detection limits for *JWST* wide- and deep-field observations are marked with a cyan dot-dashed line and a purple dashed line, respectively. The values for M_*, SFR, and $M_{\rm UV}$ corresponding to these detection limits are marked with matching styles (see Table 3.2). The axes of all quantities are oriented such that brighter, more massive, larger objects are to the top or right of the plots, and faint, low-mass, compact objects are to the bottom or left of the plots.

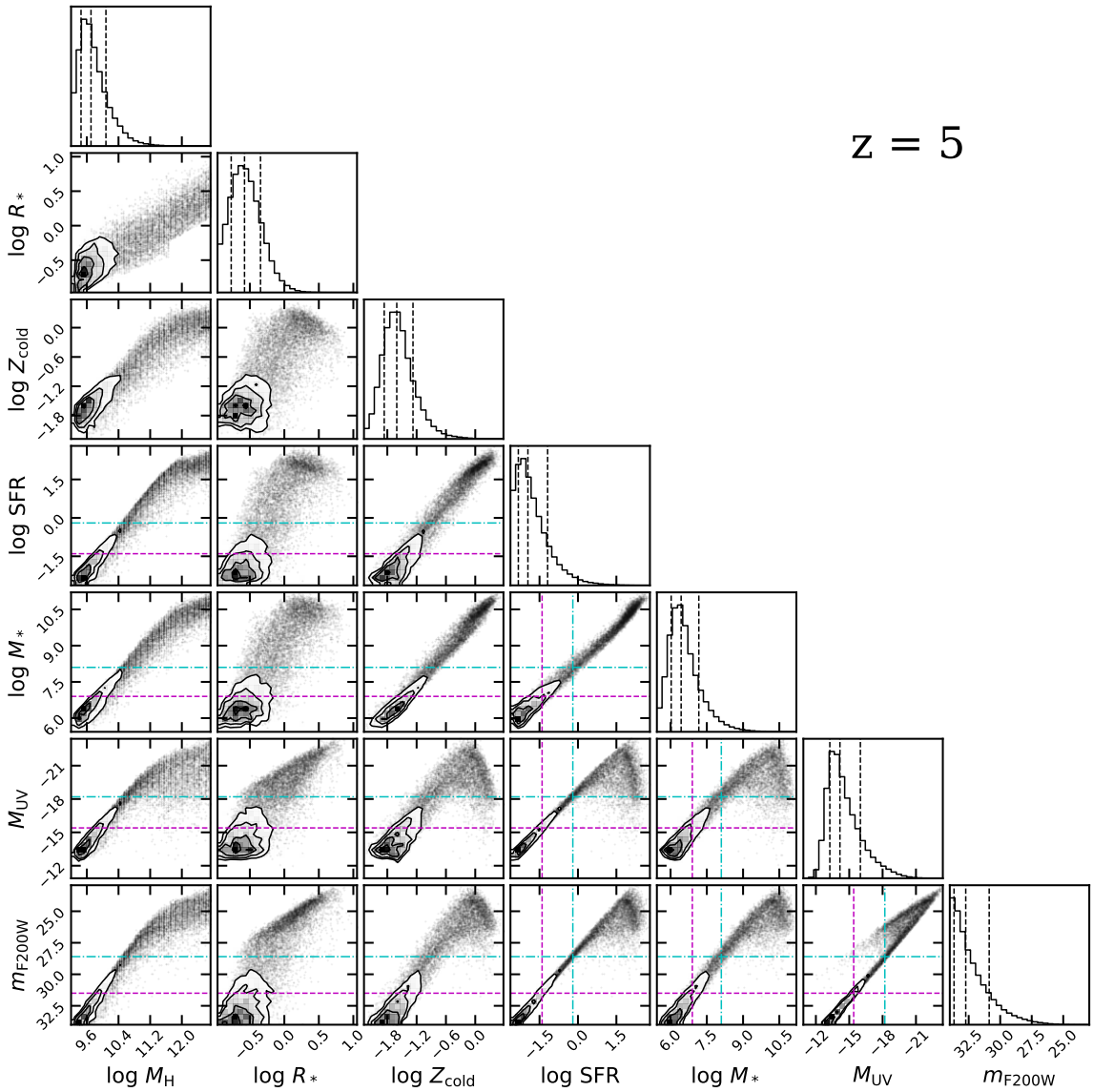

Figure 3.15 Same as Fig. 3.14 but for galaxies at $z = 5$.

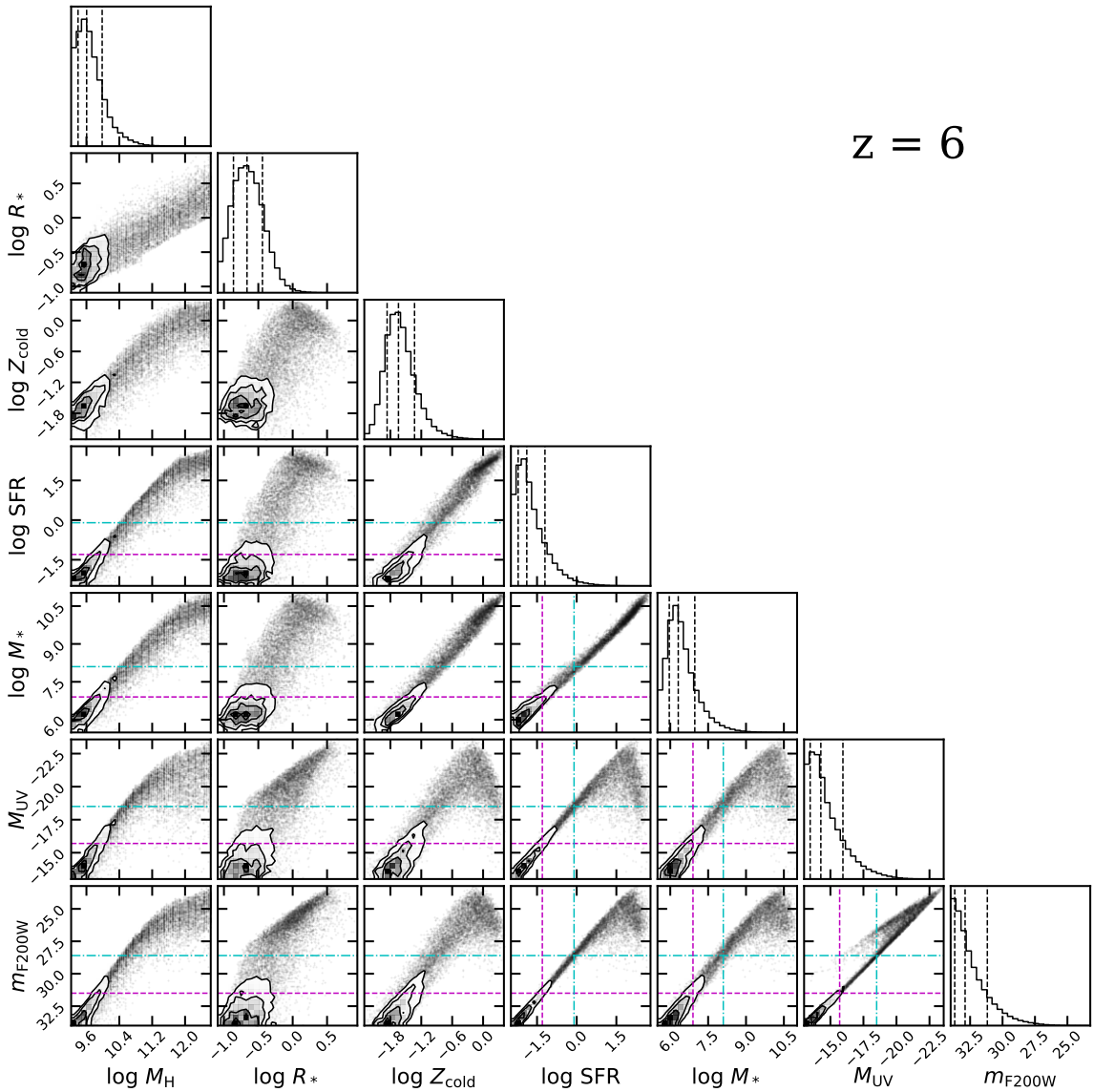

Figure 3.16 Same as Fig. 3.14 but for galaxies at $z = 6$.

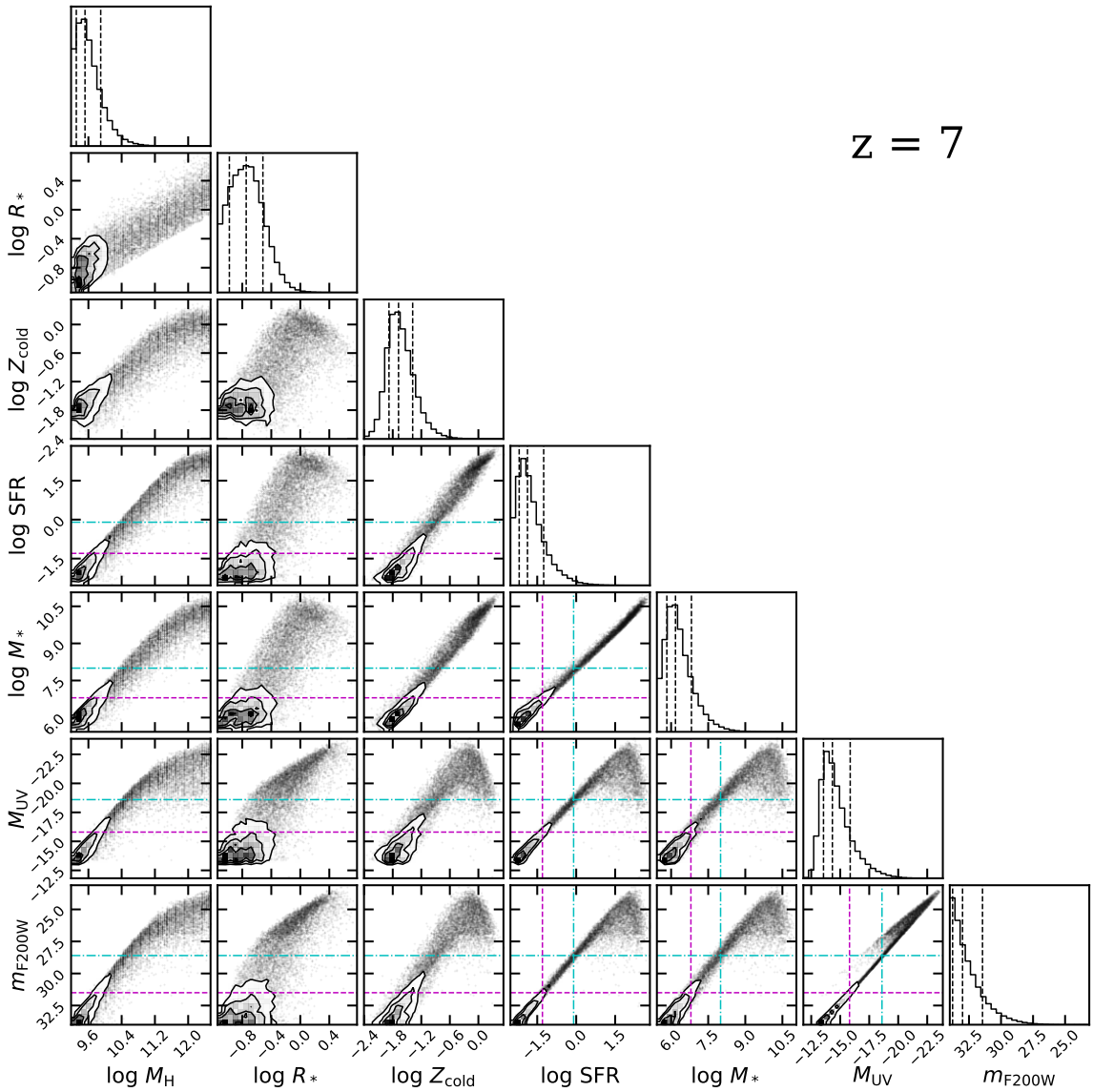

Figure 3.17 Same as Fig. 3.14 but for galaxies at $z = 7$.

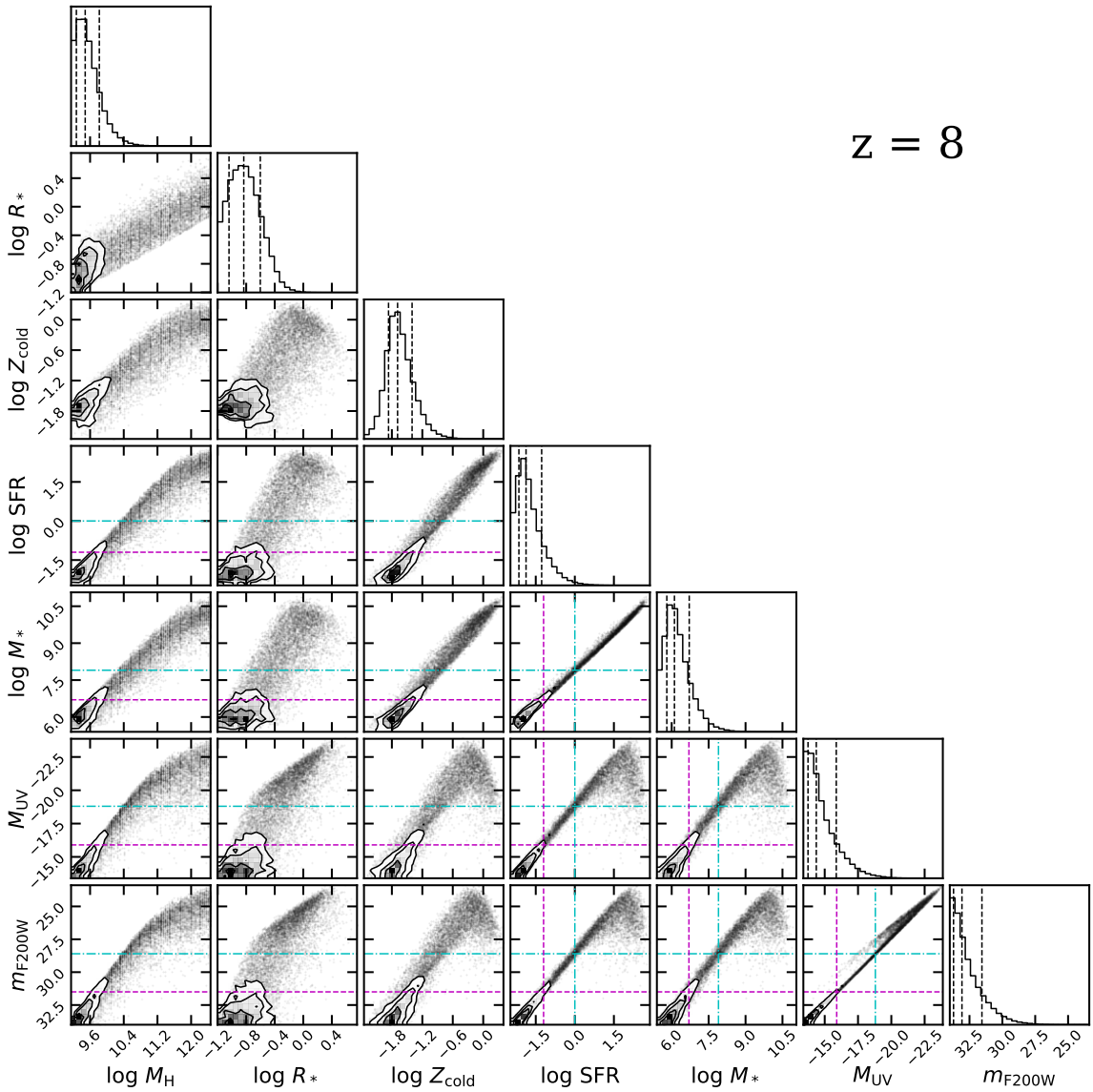

Figure 3.18 Same as Fig. 3.14 but for galaxies at $z = 8$.

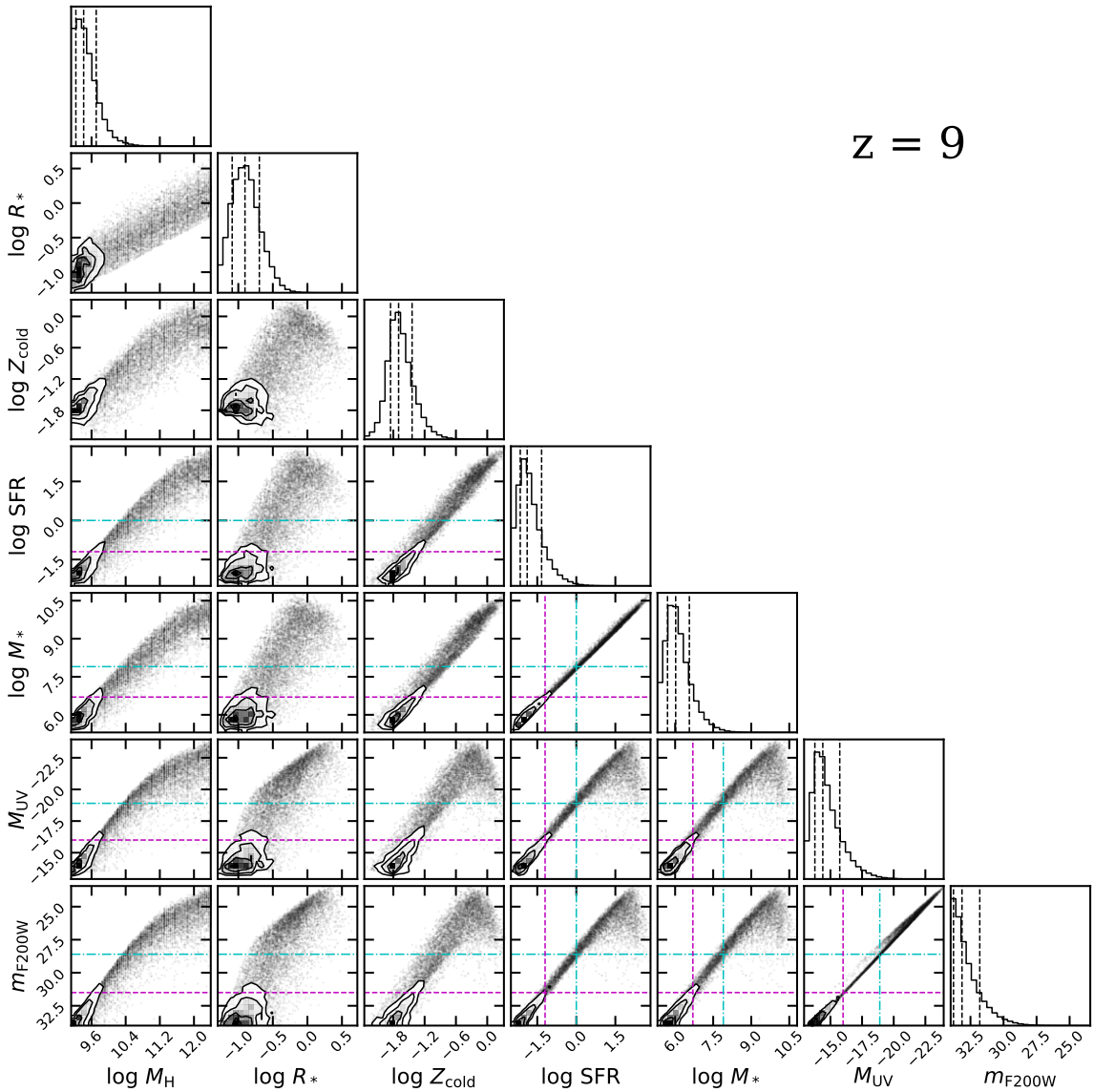

Figure 3.19 Same as Fig. 3.14 but for galaxies at $z = 9$.

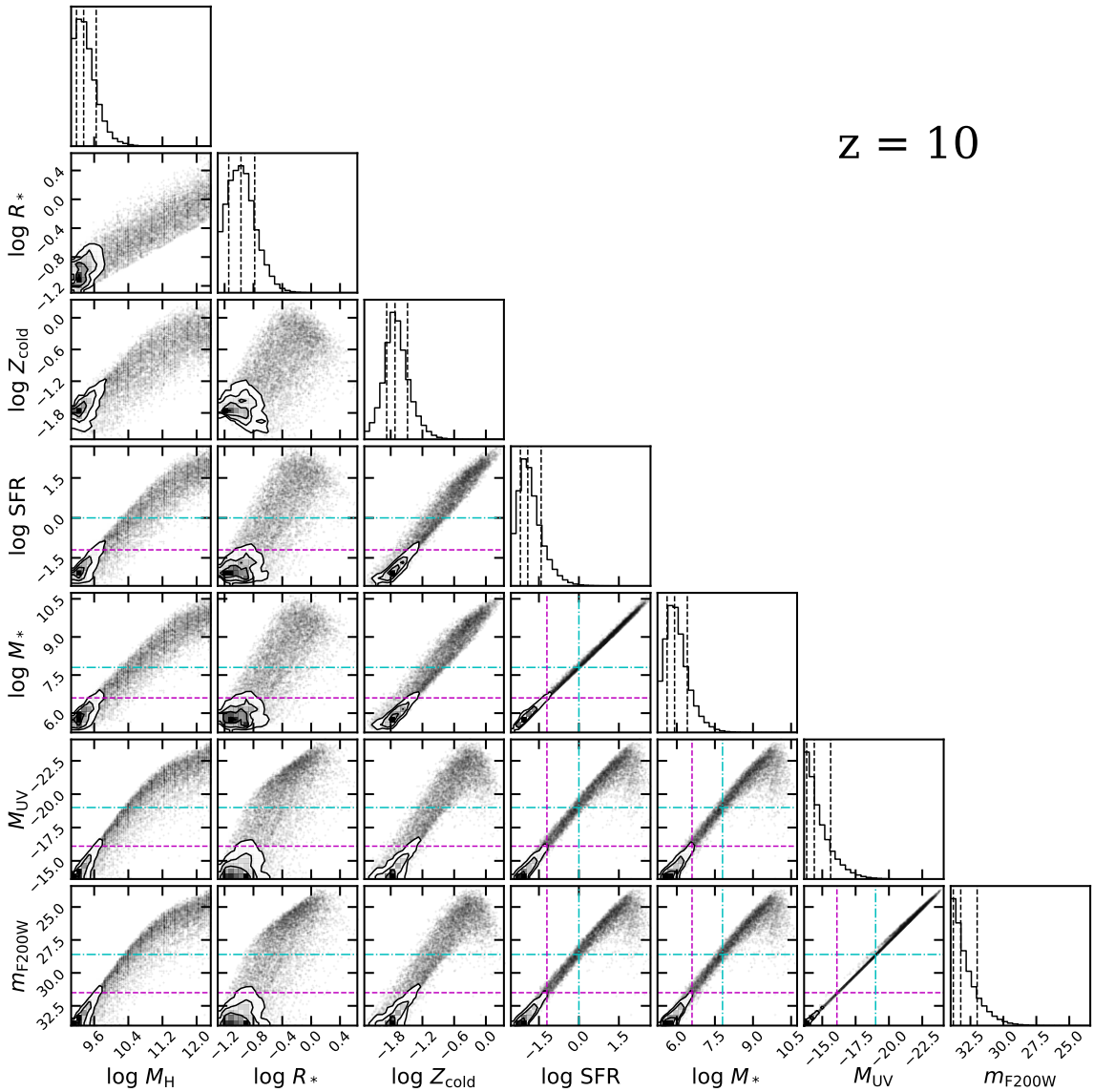

Figure 3.20 Same as Fig. 3.14 but for galaxies at $z = 10$.

the fractional difference in number density of halos in the mass range of interest is smallest at $z = 10$, and later increases rapidly towards lower redshifts. This effectively changes the underlying halo populations assumed for the predicted SHMRs and ultimately shifts the relation horizontally. Moreover, a change in cosmology leads to different accretion rates for dark matter halos, which both of these models depend on. On the other hand, as shown in Fig. 3.8, the predicted SMFs from these models seem to be in good agreement at $z \gtrsim 8$ but the predictions become quite different at $z = 4$, with the Tacchella et al. model predicting ~ 0.5 dex more galaxies than UniverseMachine. This discrepancy is a result of the differences in observational constraints at low redshift used for calibration and the assumptions made in these models (see explanation in Section 3.3.3).

In addition to the empirical models, we compare our results to the FIRE-2 simulations (Ma et al. 2018b) and to the clustering analysis by Harikane et al. (2016). Ma et al. reported both M_* and $M_{\rm h}$ for 15 galaxies at the end of the simulations ($z = 5$) and their progenitor halos at $z = 10$. The Harikane et al. analysis is based on 10,381 Lyman break galaxies (LBGs) at $z \sim 4$–7 identified in various legacy surveys. We also show the evolution of the SHMR at $z = 4$, 7, and 10, and compared that to abundance matching result at $z = 0$ from Rodríguez-Puebla et al. (2017). Our predictions for the SMHM relation are remarkably similar to the FIRE results. They are consistent with the Harikane et al. results at $z \sim 6$ and $z \sim 7$, but diverge by an increasing amount at $z \sim 5$ and $z \sim 4$. It is interesting that the clustering-based Harikane et al. results are more consistent with UniverseMachine.

In Fig. 3.22, we compare the evolution of the SHMR for a specific range of observed galaxies with $M_{\rm UV} = -21 \pm 0.25$ as presented in Finkelstein et al. (2015b). The halo masses for these observed galaxies are estimated using abundance matching with halo populations assuming the WMAP-7 cosmology. Similarly, galaxies predicted by our model and the UniverseMachine that satisfy the same criteria are selected for comparison. It is quite intriguing that abundance matching and empirical models indicate an increasing trend out to higher redshift, while our physically motivated model predicts the opposite. This highlights an interesting open question that will be addressed by future *JWST* observations.

Similarly, in Fig. 3.23, we show the evolution of the ratio between the stellar disc radius and the virial radius of the host halo (stellar-to-halo radius ratio, SRHR) for the bright ($M_{\rm UV} = -21.0 \pm 0.5$) and the faint ($M_{\rm UV} = -15.0 \pm 0.5$) populations predicted by our models. These predictions are compared to observations between $z \sim 4$–8 from Shibuya et al. (2015). Instead of showing individual measurements, we show the approximate median and uncertainties $R_*/R_{\rm H} \approx 0.019^{+0.21}_{-0.09}$. The radius shown for our model is the 3D half-stellar mass radius, which is expected to be slightly larger than the projected, rest-UV effective radius measured by observations. The rather simple relation adopted

in our model is in good agreement with the observations out to $z \sim 8$. This highlights another interesting science question that will be elucidated by *JWST* observations.

In Fig. 3.24 we show a side by side comparison of the $M_* - M_{\rm UV}$ and M_*-sSFR relation at $z = 4$, 6, 8, and 10, compared with the relations obtained from or assumed in Tacchella et al. and UniverseMachine (Behroozi et al. 2019). Using the linear fit we found for the M_*-SFR relation, we also label the x-axis with the value of SFR corresponding to the given M_*. We find that sSFR evolves quite rapidly as a function of redshift. These models agree quite well at $z = 4$ but then diverge at high redshifts. The pronounced differences at high stellar masses arise from our different treatments of dust attenuation. We can see also the different ranges in halo mass spanned by the Tacchella et al. model and UniverseMachine models, which are due to their use of numerical N-body simulations, which suffer from limited mass resolution and volume. Our predictions span the largest dynamical range, because of our use of analytic halo mass functions calibrated to multiple simulations with varying resolution and volume.

3.6 Discussion

In this section, we discuss some caveats and uncertainties in our modelling, discuss our results in the context of other results in the literature, and present an outlook for future observations beyond *JWST*.

3.6.1 Interplay between galaxy formation and cosmology

The details of when galaxies form and how quickly they evolve are somewhat sensitive to the adopted cosmology and primordial power spectrum. The estimated values of the cosmological parameters have evolved significantly throughout the past few years. For instance, the matter density parameter Ω_m measured by the Planck Collaboration (2014, 2016a) decreased significantly from previous measurement by WMAP (Komatsu et al. 2009, 2011; Hinshaw et al. 2013). This has non-trivial effects on the dark matter halo demographics (see Fig. E.1 in Appendix C). In this work, we have updated the cosmological parameters adopted in our model to the more up-to-date values reported by the Planck Collaboration in 2016, as compared to the WMAP-5 values used in previous work with these models. As Ω_m increases, the abundance of dark matter halos found in cosmological N-body simulations increases across all mass scales. In order to continue to match the same observed galaxy abundances, the relationship between galaxy luminosity or stellar mass and halo mass has to shift, such that the stellar mass occupying a halo of a given mass must decrease. In physically based models of galaxy formation such as our SAMs, this is accomplished by changing the efficiency of

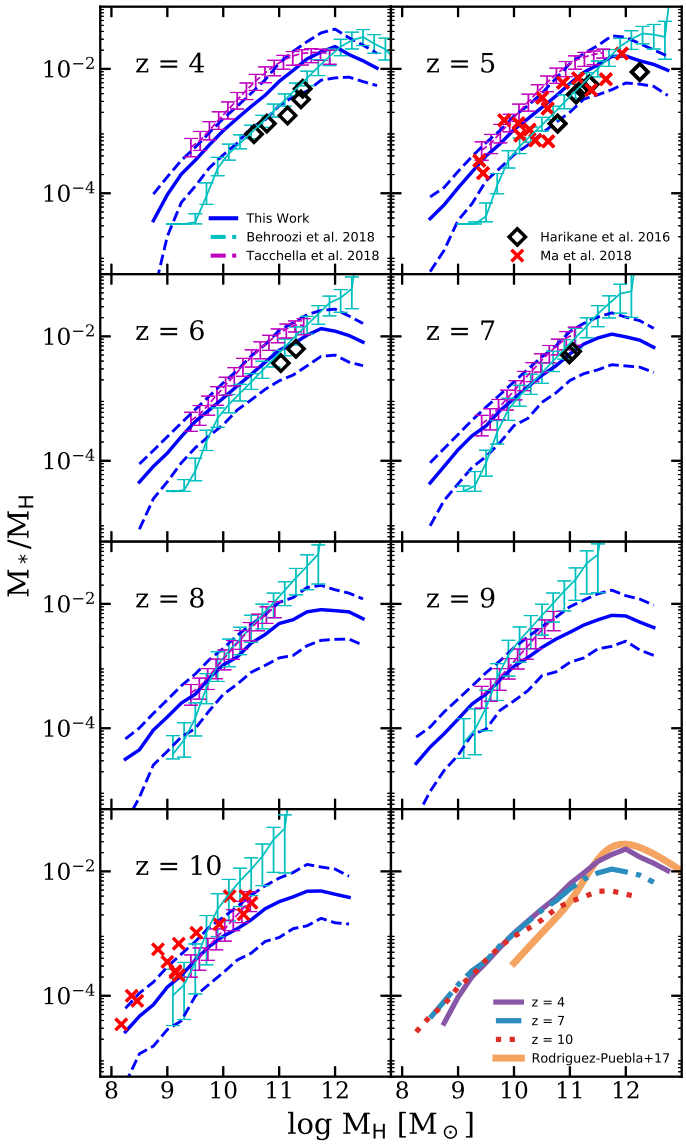

Figure 3.21 Stellar-to-halo mass ratio (SHMR) predicted by our fiducial model (dark blue). The solid line shows the median, and the dashed lines show the 16th and 84th percentiles. Our results are compared to semi-emperical models from Behroozi et al. (2019) and Tacchella et al. (2018). The error bars represent the statistical errors in these models. The red cross symbols show predictions from the FIRE-2 simulations (Ma et al. 2018b) and observational constraints from the clustering analysis by Harikane et al. (2016). The last panel show an overlay of the SHMR median predicted at $z = 4$, 7, and 10 from our model. We also show abundance matching results from Rodríguez-Puebla et al. (2017) at $z = 0$, which is used in the calibration of our model.

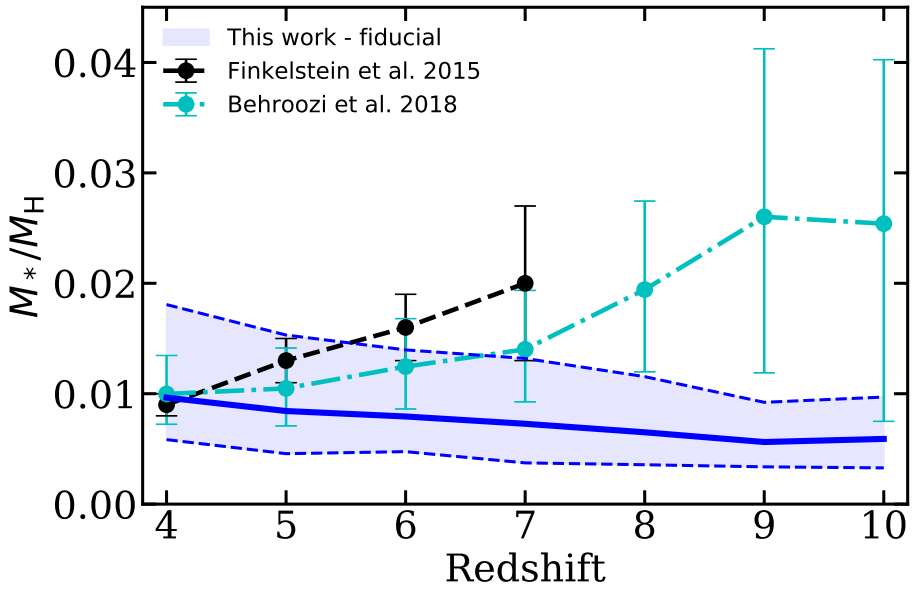

Figure 3.22 SHMR as a function of redshift predicted by our fiducial model for galaxies with $M_{\rm UV} = -21 \pm 0.25$ compared to observational estimates from Finkelstein et al. (2015b) and the empirical model UniverseMachine (Behroozi et al. 2019). The blue solid line shows the median from our model, and the dashed lines mark the 16th and 84th percentiles. The error bars for Finkelstein et al. and UniverseMachine represent observational uncertainties and statistical errors. Our models predict that the SHMR decreases slightly with increasing redshift for these galaxies, while the observational estimates and the UniverseMachine predict an increase of this quantity back in time.

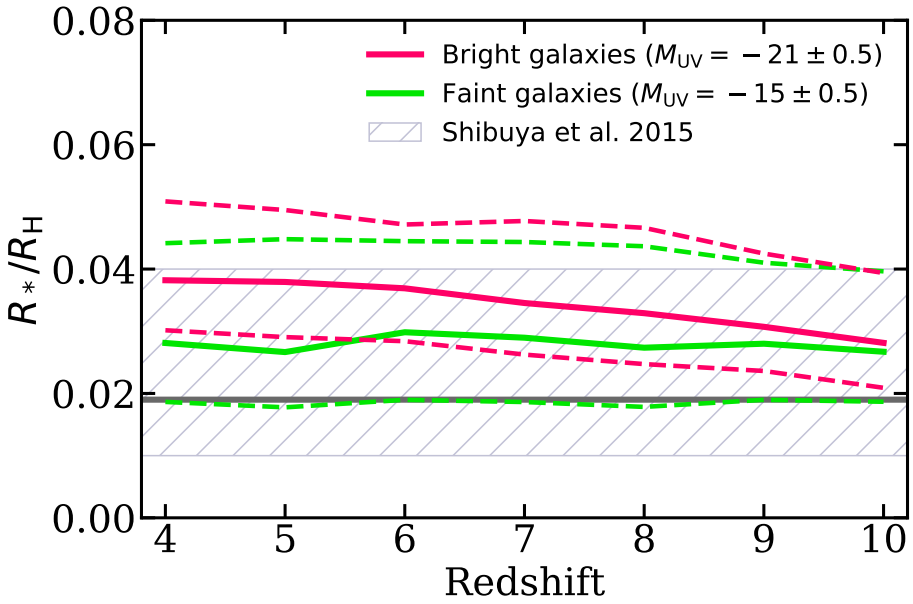

Figure 3.23 The redshift evolution of the stellar-to-halo size ratio predicted by our model, where the results for bright ($M_{\rm UV} = -21 \pm 0.5$, purple) and faint ($M_{\rm UV} = -15 \pm 0.5$, blue) galaxies are presented separately. The error bars represent the intrinsic scatter of the properties, where the upper and lower limits mark the 84th and 16th percentile, respectively. The gray band approximates the abundance matching results presented by (Shibuya et al. 2015) ($z \sim 4$–8). The radius shown for our model is the 3D half-stellar mass radius, which is expected to be slightly larger than the projected, rest-UV effective radius measured by observations. It is intriguing that the relationship predicted by our model is in good agreement with the observational constraints out to $z \sim 8$.

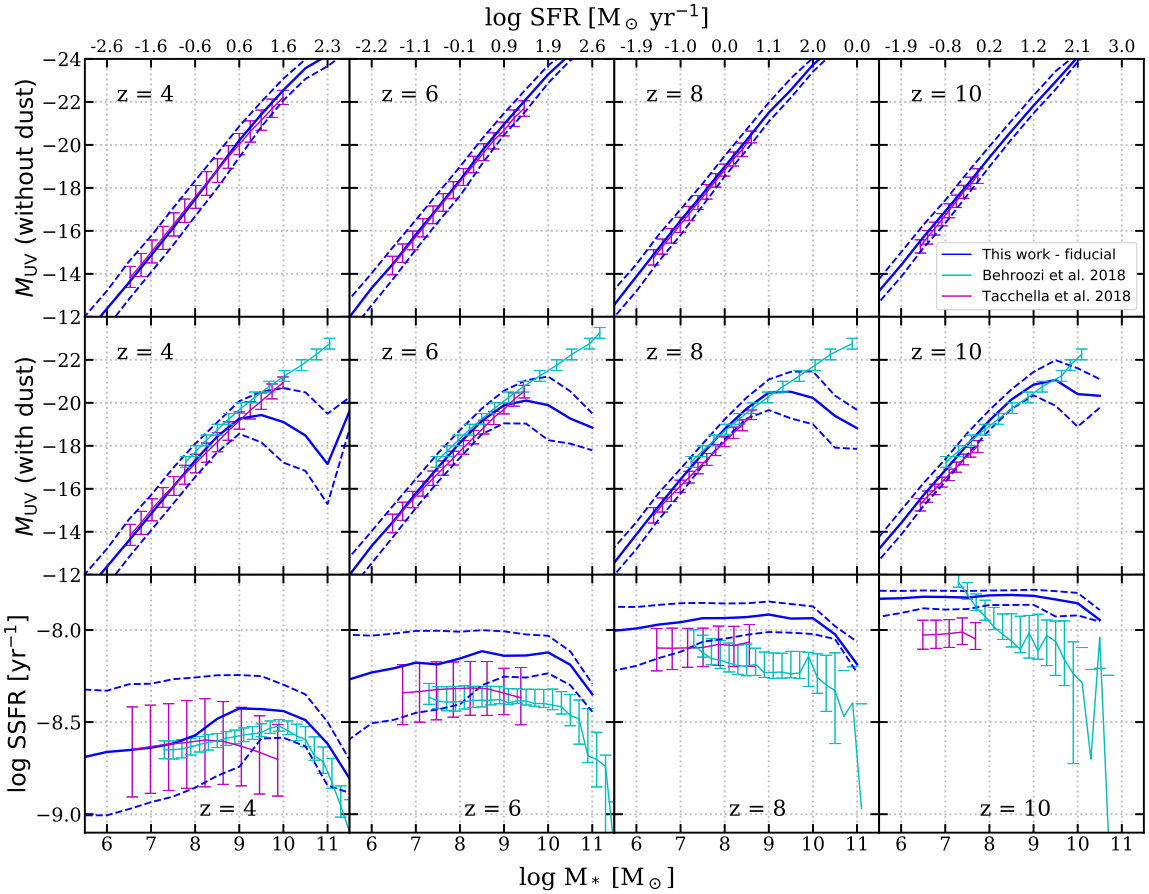

Figure 3.24 The redshift evolution of $M_{\rm UV}$ (top row: unattenuated; middle row: with dust attenuation) and sSFR predicted by our fiducial model (dark blue), where the 16th, 50th, and 84th percentiles are shown. Predictions from UniverseMachine (Behroozi et al. 2019) and Tacchella et al. (2018) are also shown in cyan and purple, respectively. The error bars represent the statistical errors on these models. This figure illustrates that there are significant differences in the assumed and predicted scaling relations both between different semi-empirical models and between semi-analytic and semi-empirical models.

stellar feedback and/or the star formation efficiency. We provide details on the changes in parameter values that were required in our models in order to recalibrate to the Planck cosmology in Chapter 2. It would seem to provide indirect support for the ΛCDM cosmology and hierarchical structure formation paradigm that implementing baryonic processes in a physically motivated manner is able to qualitatively reproduce the observed assembly of stellar mass from $z \sim 8$ to the present day.

3.6.2 Halo mass functions and merger histories

Understanding how gravitationally bound dark matter halos form and evolve over cosmic time provides the scaffolding for any cosmologically grounded model of galaxy formation. However, predictions for the global galaxy population from the reionization epoch to the present day require a very large dynamic range (at least seven orders of magnitude in halo mass). Pure dark matter plus gravity N-body simulations are commonly used to extract DM halo properties, abundances, and merger trees. However, there is currently no publicly available, self-consistent and comprehensive suite of N-body simulations that is consistent with state-of-the-art constraints on the cosmological parameters and that fully spans the required dynamic range in halo mass, resolution, and volume. Furthermore, most existing N-body simulations store only a few outputs at very early times, so that the resulting merger histories are far too coarse at high redshift. Additionally, halo finders and merger tree building algorithms have generally not been tested at these redshifts. As a result, we have made use of fitting functions calibrated to available simulation results (see Chapter 2, Appendix C), combined with merger trees created using a method based on the Extended Press Schechter (EPS) formalism. Although this approach is highly flexible and computationally efficient, allowing us to provide predictions of galaxy properties over the largest range of mass and redshift of any study yet published of which we are aware, the constructed ensembles of merger trees are not in perfect agreement with the results from N-body. We have tested our models by running them within merger trees extracted directly from N-body simulations, and find good agreement. The EPS method has the additional drawback that it does not predict spatial information or capture any dependences of halo properties and formation history on larger scale environment (which are known to exist in full N-body simulations).

3.6.3 Uncertainties in modelling of baryonic processes

In Chapter 2 of this thesis, we showed that a semi-analytic model of galaxy formation that was calibrated only with observations at $z = 0$ makes predictions for rest-UV luminosity functions at $z \sim 4$–10 that are in remarkably good agreement with available observational estimates. In this chapter, we have similarly shown that the same model also reproduces observational estimates

of the stellar mass function and star formation rate function at these epochs (within the rather large observational uncertainties). A significant uncertainty in our modelling of galaxy luminosities, particularly in the rest-UV, is our treatment of attenuation by dust, which is very simple and somewhat ad hoc. In Chapter 2, we showed that the dust attenuations adopted in our model are consistent with observational measurements of the UV slope in high-redshift galaxies, which we found encouraging. However, these measurements have very large error bars, and in addition UV slope constrains only reddening and the relationship between attenuation and reddening is not unique or well constrained, especially at these redshifts. The finding that our models also produce stellar mass and SFR distributions that are consistent with those estimated from observations is an additional consistency check, indicating that the stellar populations in high-redshift galaxies in our models are apparently similar to those inferred from fitting the multi-wavelength SEDs of real high redshift galaxies. We emphasize that the median relation, as well as the scatter in, the relationship between stellar mass and rest-UV stellar mass-to-light ratio at $z \gtrsim 4$ is still highly uncertain, and there are significant differences in the relations adopted in the literature. This is an important component in many semi-empirical models. *JWST* will greatly improve the constraints on this critical relationship. Furthermore, semi-analytic models are beginning to model the creation and destruction of dust in galaxies self-consistently (Popping, Somerville, & Galametz 2017b), which will allow for more physically robust dust modelling in future works.

One of the most striking aspects of our results, though we are not the first to demonstrate this, is that a relatively simple model qualitatively reproduces many of the fundamental properties of galaxies (such as stellar mass functions) from $z = 0$ to $z \sim 10$, without any explicit tuning to high-redshift observations or introduction of ad hoc redshift dependence in the recipes. Given the very different conditions at $z \sim 6$–10 compared to the present-day Universe, it would not have been at all surprising if our current simple phenomenological parametrizations of physical processes had broken down badly. It is also interesting to note that certain models have already been strongly ruled out by this comparison — our work, as shown in Fig. 3.4 and 3.5 (see also §3 in Chapter 2), convincingly demonstrates that in order to produce sufficient numbers of massive/luminous, high-redshift galaxies, it is necessary to adopt a star formation relation in which SFR density scales super-linearly with molecular gas surface density in dense gas. Because high-redshift galaxies tend to be more compact and have more gas at high surface density, this leads to an effective decrease in the gas depletion time (increase in star formation efficiency) at high redshift. This appears to be in qualitative agreement with existing observations of cold gas at high redshift (Obreschkow & Rawlings 2009a; Dutton et al. 2010; Dutton & van den Bosch 2012; Saintonge et al. 2013; Decarli et al. 2016; Schinnerer et al. 2016; Krogager et al. 2018; Tacconi et al. 2018), but future observations

will provide improved constraints.

Although the fundamental physics operating in our Universe should remain unchanged across space and time, changes in the prevalent physical conditions could lead to effective evolution in redshift or cosmic time for the processes that shape galaxies. For example, processes that involve interaction with the cosmic environment, such as dark matter halo mergers, photoionization squelching, or cosmological accretion, depend on the background density or temperature, which is redshift dependent. Local processes, such as AGN and stellar feedback, are generally parametrized as a function of a galaxy properties that evolve across redshift and thus gain de facto dependency on redshift. Since each of these processes operates under different conditions and depends on a different set of properties, in principle one might be able to break certain degeneracies by studying a wide range of galaxy types over different snapshots in cosmic time. However, the baryonic processes in SAMs are parametrized in an extremely simple and phenomenological manner, and it is unclear whether these parametrizations will properly capture these multivariate correlations. Further close comparisons of SAM predictions with those from numerical hydrodynamic simulations are important to validate this approach.

There are also physical processes that may be important at extreme redshifts that are not included in our models. For example, our models do not directly model Pop III stars, nor metal enrichment by these objects, but instead assume that all top-level halos are polluted up to a metallicity floor $Z_{\text{pre-enrich}} = 10^{-3} Z_{\odot}$. Our models do not include self-consistent modelling of photo-ionization feedback ('squelching') by a meta-galactic ionizing background, but instead assume that reionization occurs everywhere in the Universe at a fixed redshift of $z \sim 8$. As discussed in Chapter 2, in some past works, squelching was thought to have a significant impact on galaxy formation in halos up to masses of $\sim 10^{10}$ $\mathrm{M}_{\odot}$ (e.g. Efstathiou 1992; Bullock, Kravtsov, & Weinberg 2000; Gnedin 2000; Somerville 2002). However, in more recent studies, the halo mass where squelching has a significant impact has dropped to much lower masses. In this work, we adopted the characteristic mass fitting function from Okamoto et al. (2008) and found that squelching has a negligible effect on observable galaxies in the mass and redshift range that we studied. This is in agreement with results from the Cosmic Reionization On Computers (CROC, Gnedin & Kaurov 2014) simulations, but not with the Cosmic Dawn simulation (CoDa, Ocvirk et al. 2016), which found that the photoionizing background has a strong effect on halos with a mass at $z \sim 3$ of $M_{\mathrm{H}} \lesssim 10^{10} \mathrm{M}_{\odot}$. The interplay between photo-ionization feedback and other feedback processes such as stellar feedback is extremely complex and must be further investigated using simulations with self-consistent radiative transfer (e.g. Finlator et al. 2011, 2018).

Another set of processes that is highly uncertain in our models is the seeding and growth of, and

feedback from, supermassive black holes. In our current models, we do not include the radiation from AGN in the galaxy SED that we compute, nor do we believe our current treatment of AGN feedback at high redshift to be realistic. This could have an important effect on galaxy properties and reionization, and will be the topic of future work.

3.6.4 Our results in the context of other model predictions

We have performed a fairly comprehensive comparison of our predictions for galaxy stellar mass functions and SFR functions at $z \sim 4$–10 with available predictions from the literature based on the three major existing techniques: semi-analytic models, numerical hydrodynamic simulations, and semi-empirical and empirical models. Somewhat surprisingly, we find that predictions from semi-analytic models and numerical hydrodynamic simulations from several different groups are in generally very good agreement for these basic quantities. This is surprising because these calculations have been done using different codes, different sub-grid treatments of physical processes, and at different resolutions. We see a general consistency among models based on *a priori* modelling of physical processes within a ΛCDM cosmological framework. Semi-empirical and purely empirical models show larger dispersions, unsurprisingly, in the regions where observational constraints are currently unavailable. This see alsoms to be encouraging news for efforts to use physically based models for forecasting and planning for future observations.

One of the fundamental questions in galaxy formation and cosmology is how galaxy properties are related to the underlying dark-matter dominated mass distribution. A simplified form of this relationship is often presented in terms of the relationship between galaxy stellar mass and the mass of its host dark matter halo (stellar-mass-halo-mass relation). Our models directly predict this relationship, as well as the dispersion in it, and we find the interesting result that our models predict almost no evolution in the median $M_*/M_{\rm H}$ from $z \sim 10$ to $z \sim 4$ for low-mass halos ($M_{\rm H} \lesssim 10^{11}{\rm M}_\odot$), while they predict almost an order of magnitude increase over this interval in the most massive halos of these epochs ($M_{\rm H} \sim 10^{12}{\rm M}_\odot$).

We compared our predictions for $M_*/M_{\rm H}$ with the results of semi-empirical models from Tacchella et al. and UniverseMachine at high redshifts. We find significant differences both *between* these two published semi-empirical models and our model predictions. Some of these differences may be due to differences in the assumed underlying cosmology and the observations used to derive the semi-empirical models. Perhaps most strikingly, both Tacchella et al. and UniverseMachine predict that $M_*/M_{\rm H}$ continues to monotonically increase with increasing halo mass at $z \gtrsim 6$, while our models predict that $M_*/M_{\rm H}$ turns over at high halo masses. These massive halos are exceedingly rare, and no numerical simulations that we are aware of have investigated this.

Our results are in tension with existing observational estimates of the evolution of $M_*/M_{\rm H}$ over the interval $z \sim 4$–7 from Finkelstein et al. (2015b) for a stacked sample of luminous galaxies ($M_{\rm UV} \sim -21$), however, we again emphasize the currently very large uncertainties on these observational estimates. *JWST* is unlikely to place strong constraints on the abundances of these extremely luminous high redshift galaxies, due to its small field of view and limited lifetime, but future wide deep surveys with instruments such as WFIRST can be anticipated to do so. *JWST* will however be able to obtain improved estimates of the redshift and stellar populations in luminous high redshift galaxy candidates that have already been discovered through *HST* and Spitzer.

If it is really true that $M_*/M_{\rm H}$ remains essentially constant from $z \sim 4$–10 in the halos that will dominate the populations observed by *JWST* (as our models predict), this has an interesting implication. It implies that the build-up in galaxy number density over this period is driven by the evolution in the dark matter halo mass function. Therefore, if we can somehow observationally constrain this $M_*/M_{\rm H}$ relationship, *JWST* observations of high redshift galaxies could provide interesting constraints on cosmology. Another interesting point is that, as can be clearly seen in Fig. 3.2 (or see also Fig. 13 in Chapter 2), the expected turnover in the stellar mass or luminosity function from baryonic processes is well below the anticipated sensitivity of *JWST* even for lensed fields. Therefore, if a turnover or cutoff at faint magnitudes is seen, it could be a sign of a cutoff in the small scale power spectrum, such as that expected in certain exotic varieties of dark matter.

3.6.5 Outlook for future observations with *JWST* and beyond

One of the main focuses of this work is to establish the connections between the predicted rest-frame UV luminosities and observed-frame IR magnitudes for high-redshift galaxies (presented in Chapter 2) to their intrinsic physical properties. From the wide range of physical properties predicted by our SAM, we have selected a few that are of the greatest interest, including stellar mass M_*, SFR, cold gas phase metallicity $Z_{\rm cold}$, and stellar radial size R_*. A comprehensive view of the correlations among these properties along with $m_{\rm F200W}$, $M_{\rm UV}$, and halo mass $M_{\rm H}$ have been shown in Fig. 3.16, 3.18, and 3.20. These corner plots can be used as lookup tables to facilitate both the planning of observations and simulations. For instance, simulators can use the first column to quickly estimate the range of galaxy properties expected for some given halo mass, and conversely, observers can use the bottom row to estimated the physical properties of an observed population given an object's observed-frame IR magnitude.

We applied selection criteria for representative 'wide', 'deep' and 'lensed' *JWST* surveys to our model predictions, in order to predict the range of stellar masses and SFR that will be probed. By combining these 'sensitivity' functions with the experiments that we carried out in Section 3.3.2,

in which we varied the parameters controlling various physical processes in the models, we can see what kind of survey design characteristics will be necessary to probe different physical processes that are currently highly uncertain in galaxy formation models. For example, we can clearly see that the planned wide and deep surveys with *JWST* will help to constrain stellar feedback processes at high redshift. Constraining the efficiency of star formation at redshifts greater than six will likely require larger area surveys than will be feasible with *JWST*.

Apart from the planned *JWST* GTO and ERS programs mentioned above, there will be plenty of other observational opportunities to probe the very early Universe with upcoming facilities, including the space-based Euclid (Racca et al. 2016) and Wide-Field Infrared Survey Telescope (WFIRST, Spergel et al. 2015), as well as the ground-based Large Synoptic Survey Telescope (LSST Science Collaboration 2017). Future deep Atacama Large Millimeter Array (ALMA) surveys will be able to put constraints on the cold gas mass in $z \sim 3$–7 galaxies via CO and dust continuum observations, as well as to probe the ISM conditions through fine-structure lines.

Beyond individual source detection, intensity mapping is a new technique being developed to indirectly constrain the high-redshift galaxy population (Visbal & Loeb 2010; Visbal, Trac, & Loeb 2011; Kovetz et al. 2017) over large areas of the sky. Numerous on-going intensity mapping experiments for HI, CO, CII, and Lyα are planned or underway, including BINGO (Battye et al. 2013), CHIME (Bandura et al. 2014), EXCLAIM (Padmanabhan 2019), HERA (DeBoer et al. 2017), HIRAX (Newburgh et al. 2016), Tianlai (Chen 2012), LOFAR (Patil et al. 2017), MeerKat (Pourtsidou 2018; Santos et al. 2017), CONCERTO (Serra, Doré, & Lagache 2016), PAPER (Parsons et al. 2010). The development of efficient and robust modelling techniques to interpret results from all of these upcoming experiments will be critical to realize their full scientific potential.

3.7 Summary and Conclusions

In this work, we presented predictions from semi-analytic models for physical properties of galaxies at $z = 4$–10, an epoch that will be probed by upcoming observations with *JWST*. With EPS-based merger trees, we can efficiently sample halos over a wide mass range, ranging from the ones near the atomic cooling limit to the most massive ones at a given redshift, unlike numerical simulations which are much more limited in the dynamic range that can be simulated. We showed distribution functions for statistical properties such as stellar mass and SFR functions, and investigated how varying the parameters in recipes for the physical processes that shape these galaxy populations affect these results. In addition, we provide predictions of the scaling relations between physical properties that are directly predicted by our simulations, as well as between intrinsic and observable

properties. We provide predictions for the properties of the galaxy populations that will be probed by representative wide, deep, and lensed *JWST* surveys.

We have shown the one-point distribution functions for stellar mass, star formation rate, cold gas mass, and molecular gas mass between $z = 4$–10 predicted by our semi-analytic model. We have also studied and quantified the impact of uncertainties in our parametrizations of key processes such as star formation efficiency and the mass loading of stellar driven winds. The free parameters in our model are calibrated once to a subset of observational constraints at $z \sim 0$ and are not retuned to match observations at higher redshifts. Thus, it is encouraging that the predictions produced by our fiducial model are in good agreement with observational constraints for SMFs and SFRFs up to $z \sim 8$, although the uncertainties on the observational estimates of these quantities are currently quite large. This suggests that the approach used to model physical processes in galaxy formation models, which has been quite successful at lower redshift, is not failing badly even at these very early epochs. We also compare the predictions for SMF and SFRFs from our semi-analytic approach with those from numerical hydrodynamic simulations and semi-empirical models. We find reasonable agreement, particularly with the numerical simulations, reinforcing the robustness of these model predictions.

The scaling relations of a range of galaxy properties and their evolution are also studied and compared to other work in the literature. For example, we present the relationships between halo mass, stellar mass, galaxy radius, cold gas phase metallicity, SFR, and rest-frame and observed luminosity (with and without dust attenuation). We present all of these relationships at $z = 6$, $z = 8$, and $z = 10$, and provide extensive tables and fitting formulae describing the results. We hope that these predictions will be helpful both for interpreting future observations as well as for planning high-resolution zoom-in simulations.

By forward modelling the predicted star formation and chemical enrichment histories into the observed-frame IR magnitude for *JWST* NIRCam filters, we are able to select galaxies according to criteria intended to mimic representative wide, deep, and lensed *JWST* surveys. We show the predicted distributions of physical properties for objects that would be detected in these surveys. We found that wide- and deep-field surveys will be able to probe galaxies down to $M_* \sim 10^8 \mathrm{M}_\odot$ and $\sim 10^7 \mathrm{M}_\odot$ at $z = 4$, respectively. Our model also predicts that high-redshift galaxies are intrinsically brighter than their lower redshift counterparts of similar mass. Thus, the M_* corresponding to the detection limit evolves rather mildly as a function of redshift (see Table 3.2).

Taking advantage of our model's efficiency, we are able to quantify some of the effects of the uncertainties in the empirical physical recipes on the resultant galaxy population by systematically varying several key model parameters. We find that the gas depletion time and star formation

relation have degenerate effects on the massive galaxy populations, while stellar feedback strongly influences star formation in low-mass halos. On the other hand, the current implementation of photoionization feedback from a metagalactic UV background, as well as AGN feedback, do not have a strong impact on our model predictions at these epochs.

We summarize our main conclusions below.

1. The predictions from our fiducial model, which was calibrated only at $z = 0$, are consistent with observed SMFs and SFRFs between $z \sim 4$–8. This suggests that our implementation of the physical processes that shape galaxy populations may still be reasonably accurate at these very early times.

2. High-redshift SMFs from *JWST* will help constrain our physical understanding of the build-up of galaxies. At high masses, the mass-function shape is significantly affected by the star-formation efficiency or time-scale. At the low-mass end it is most influenced by the modelling of stellar-driven outflows.

3. We find that high-redshift galaxies are intrinsically brighter than their lower-redshift counterparts of similar stellar masses due to their higher SFR and younger stellar populations. In addition to that, massive galaxies at high redshifts have higher dust-extincted luminosities due to lower dust content. This results in a fixed *observed-frame* m_{F200W} selecting populations with similar stellar mass limits across the redshift range $z = 4$–10.

4. We find that changing the slope of the scaling of the mass loading factor of stellar-driven outflows with galaxy circular velocity, α_{rh}, changes both the slope of the low-mass end of the SMF and the mass where the SMF and LF cut off due to inefficient cooling. Changing α_{rh} from a value of 2.0 to 3.6 results in a change in the number density of the lowest mass galaxies ($M_* \lesssim 10^8$ M$_\odot$) of 0.5–1 dex. We find that changing the star formation efficiency by a factor of two has a negligible impact on the low mass end of the SMF, but changes the number density of massive galaxies by ~ 0.1–~ 0.25 dex.

5. There is remarkably good agreement between the predicted SMF from theoretical simulations based on different techniques and codes, in particular between our semi-analytic models and numerical hydrodynamical simulations. Larger discrepancies (up to nearly 1 dex for the lowest mass galaxies, and also for very massive galaxies) are seen between models when semi-empirical models are included in the comparison. Future observations with *JWST* will be able to help discriminate between these models.

6. Anticipated wide-field (deep-field) *JWST* NIRCam surveys will be able to probe galaxies with rest-frame $M_{\rm UV} \sim -17.28$ (-14.72) at $z \sim 4$ and -18.73 (-16.06) at $z \sim 10$, which corresponds to $\log M_*/{\rm M}_\odot \sim 7.96$ (6.78) at $z \sim 4$ and 7.51 (6.35) at $z \sim 10$ and $\log {\rm SFR}/({\rm M}_\odot\ {\rm yr}^{-1}) \sim -0.4$ (-1.54) at $z \sim 4$ and -0.10 (-1.24) at $z \sim 10$.

7. None of models considered in this chapter predict low-mass truncations or turnovers in the SMF at fluxes bright enough to be observed by *JWST*, even in lensed fields.

Chapter 4

Intrinsic Production Efficiency of Lyman-Continuum Radiation

The *James Webb Space Telescope* (*JWST*) is expected to enable transformational progress in studying galaxy populations in the very early Universe, during the Epoch of Reionization (EoR). A critical parameter for understanding the sources that reionized the Universe is the Lyman-continuum production efficiency, $\xi_{\rm ion}$, defined as the rate of production of ionizing photons divided by the intrinsic UV luminosity. In this work, we combine self-consistent star formation and chemical enrichment histories predicted by semi-analytic models of galaxy formation with stellar population synthesis (SPS) models to predict the expected dependence of $\xi_{\rm ion}$ on galaxy properties and cosmic epoch from $z = 4$–10. We then explore the sensitivity of the production rate of ionizing photons, $\dot{N}_{\rm ion}$, to the choice of SPS model and the treatment of stellar feedback in our galaxy formation model. We compare our results to those of other simulations, constraints from empirical models, and observations. We find that adopting SPS models that include binary stars predict about a factor of two more ionizing radiation than models that only assume single stellar populations. We find that UV-faint, low-mass galaxies have values of $\xi_{\rm ion}$ about 0.25 dex higher than those of more massive galaxies, but find weak evolution with cosmic time, about 0.2 dex from $z \sim 12$–4 at fixed rest-UV luminosity. We provide predictions of $\dot{N}_{\rm ion}$ as a function of $M_{\rm h}$ and a number of other galaxy properties.

This chapter is reproduced from published work *Semi-analytic forecasts for JWST – III. Intrinsic production efficiency of Lyman-continuum radiation* (Yung et al. 2020. MNRAS, 494, 1002).

4.1 Introduction

The ionization of hydrogen in the intergalactic medium (IGM) is a critical landmark in cosmic history. The onset and duration of this transition is constrained by various kinds of observations, including the polarization of the cosmic microwave background (CMB), the spectra of high-redshift ($z \gtrsim 6$) quasars, and the abundance of Lyα emitters (e.g. Fan et al. 2006a; Robertson et al. 2013). Despite the uncertainties in these observational constraints, the astronomical community has reached a broad consensus that the process of hydrogen reionization took place roughly between $z \approx 6$–10,

which is commonly referred to as the Epoch of Reionization (EoR) (Loeb & Barkana 2001). However, there has been ongoing tension between the total ionizing photon budget accounted for by all known sources and the time frame of cosmic reionization set by current observations.

Although it is clear that the population of directly observed high-redshift galaxies observed to date alone are insufficient to ionize the IGM (e.g. Bouwens et al. 2015a; Finkelstein et al. 2015a; Robertson et al. 2015), many recent analytic studies have shown that a reasonable extrapolation of the observed galaxy populations during the EoR down to fainter rest-frame UV luminosities (e.g. $M_{\rm UV} \sim -13$ may be able to account for most reionization constraints (Finkelstein et al. 2012a, 2015a, 2019; Kuhlen & Faucher-Giguère 2012; Bouwens et al. 2015a; Robertson et al. 2015; Naidu et al. 2020). This is further supported by cosmological hydrodynamic simulations, many of which are able to satisfy the constraints (e.g. Geil et al. 2016; Ocvirk et al. 2016, 2020; Anderson et al. 2017). However, there are still quite a few remaining uncertainties in the estimation of the production rate of ionizing photons.

The total number of ionizing photons available to reionize the IGM depends on the abundance of star-forming galaxies during the EoR, their intrinsic production efficiency of Lyman-continuum (LyC) radiation, and the fraction of radiation that escapes to the IGM. Each of these three moving parts comes with its own substantial set of uncertainties in both modelling and observing, and are extremely challenging to model self-consistently since they involve physical processes that span across many orders of magnitude in physical scales.

The intrinsic production rate of ionizing photons, $\dot{N}_{\rm ion}$, is simply the number of ionizing photons being produced per second by a galaxy, which can be formally obtained by directly integrating over the galaxy SED

$$\dot{N}_{\rm ion} = \int_{\nu_{912}}^{\infty} L_\nu (h\nu)^{-1}\, d\nu. \tag{4.1}$$

This quantity is often normalized to the intrinsic rest-frame far ultraviolet (FUV) luminosity and expressed as the ionizing photon production efficiency

$$\xi_{\rm ion} = \dot{N}_{\rm ion} \,/\, L_{\rm FUV}, \tag{4.2}$$

where note that $L_{\rm FUV}$ is the intrinsic UV luminosity, not including the effects of attenuation by dust, and is therefore in general not directly observable. The expected integrated value of $\dot{N}_{\rm ion}$ or $\xi_{\rm ion}$ for a galaxy therefore depends on its spectral energy distribution (SED), which depends on the initial mass function (IMF) and joint distribution of ages and metallicities of the stellar populations contained in it, as well as on the physics of stellar evolution and stellar atmospheres.

Decades of effort have been invested into modelling the SED arising from 'simple' (single age and metallicity) stellar populations (SPS models). Examples include STARBURST99 (Leitherer et al. 1999), PÉGASE (Fioc & Rocca-Volmerange 1997, 1999, 2019), Bruzual & Charlot (2003, hereafter BC03, see also Bruzual 1983; Bruzual & Charlot 1993; Charlot & Bruzual 1991), Maraston (2005), Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009, 2010; Conroy & Gunn 2010), Binary Population and Spectral Synthesis (BPASS; Stanway et al. 2016; Eldridge et al. 2017; Stanway & Eldridge 2018). These models combine stellar isochrones with stellar atmosphere models or templates, weight them with an assumed stellar initial mass function, and provide stellar continuum SEDs for a grid of stellar ages and metallicities. Although the predictions of SPS models are fairly robust and well converged at optical wavelengths, the modelling of the massive stars that produce ionizing radiation suffers from much larger uncertainties, including the proper treatment of convection, rotation, and the evolution of binary stars. Moreover, it is unknown whether the stellar IMF during the EoR resembles the one in nearby galaxies, or whether the IMF has a significant dependence on the environment in which stars are born, which might lead to an indirect dependence on redshift or global galaxy properties.

In the past, coarse estimates of $\xi_{\rm ion}$ have been obtained indirectly using the observed UV-continuum slope, $\beta_{\rm UV}$, which is used as a proxy for predominant massive, UV-bright stars (Robertson et al. 2013; Bouwens et al. 2015a, 2016a; Duncan & Conselice 2015) or using the measured stellar age and metallicities (e.g. Madau et al. 1999; Schaerer 2003). Recently, measurements of nebular emission lines from spectroscopy have enabled more direct constraints on $\xi_{\rm ion}$; both of which are used in conjunction with stellar population synthesis models. Recent studies have attempted to use measurement of nebular emission lines and UV-continuum fluxes to more directly infer the total number of LyC photons produced and put constraints on $\xi_{\rm ion}$. Bouwens et al. (2016c) and Lam et al. (2019) provided estimates with Hα at $z \sim 4$–5, and Stark et al. (2015) used CIV as a tracer for a galaxy at $z \sim 7$. The highly anticipated Near Infrared Spectrograph (NIRSpec) onboard the *James Webb Space Telescope* (*JWST*), as well as future Extremely Large Telescope (ELT) observations are expected to obtain high-resolution spectra for more robust determination of many physical properties including $\xi_{\rm ion}$. Moreover, low-redshift galaxies ($z \lesssim 2$) are sometimes viewed as analogues of their high-redshift counterparts, which may provide complementary insights into the otherwise hard-to-measure quantities (Nakajima et al. 2016, 2018; Schaerer et al. 2016; Matthee et al. 2017; Shivaei et al. 2018).

Many previous analytic calculations of reionization have assumed a constant value for $\xi_{\rm ion}$ (e.g. Madau et al. 1999; Robertson et al. 2015) or adopted a simple parametrization as a function of redshift (e.g. Finkelstein et al. 2012a). Kuhlen & Faucher-Giguère (2012) attempted to explore the

uncertainties in $\xi_{\rm ion}$ due to lack of knowledge about the stellar populations and SED in their models by parametrizing it and exploring a bracketing range of values. Finkelstein et al. (2019, hereafter F19) parametrize $\xi_{\rm ion}$ as a function of $M_{\rm UV}$ and redshift and allow these parameters to be fit as part of a multi-parameter Bayesian Monte Carlo Maximum Likelihood fitting procedure. Many (semi-)numerical simulations that aim to capture the evolution of large-scale structure, but do not contain detailed modelling of galaxy formation, assume simple scaling relations that allow $\xi_{\rm ion}$ or $\dot{N}_{\rm ion}$ to scale with stellar mass or halo mass (Mesinger & Furlanetto 2007; Choudhury et al. 2009; Choudhury 2009; Santos et al. 2010; Hassan et al. 2016, 2017).

$\dot{N}_{\rm ion}$ and $\xi_{\rm ion}$ for composite stellar populations depend on the joint distribution of ages and metallicities in each galaxy, and the values will be highly sensitive to the burstiness of the star formation history. These quantities may in principle evolve with redshift or be correlated with global galaxy properties such as mass or luminosity. It is thus important to use a self-consistent, physically grounded approach that hopefully contains the relevant internal correlations between physical properties. Cosmological models that follow the formation of stars and chemical evolution in galaxies can, in principle, straightforwardly obtain more physically motivated estimates of $\xi_{\rm ion}$ by simply convolving their predicted star formation and chemical evolution histories with SPS models like those described above. Wilkins et al. (2016b, hereafter W16) did exactly this using the large volume BlueTides simulations, and presented predictions for how $\xi_{\rm ion}$ evolves with redshift and its dependence on stellar mass. Similar work has also been done with the FIRE (Ma et al. 2016b), Vulcan (Anderson et al. 2017), and FirstLight zoom-in simulations (Ceverino et al. 2019).

In this chapter, we focus on understanding the production efficiency of ionizing photons in EoR galaxies, which is a quantity that will be able to constrained with future *JWST* photometric and spectroscopic surveys. Based on the results from the well-established physical models that have been rigorously tested in previous works, we explore the scaling relations between the intrinsic production efficiency of ionizing photons and a wide range of SF-related physical properties, as well as other observable properties. These results are also compared to predictions from other simulations and values inferred by existing observations. Our computationally efficient semi-analytic approach allows us to span a larger dynamic range in galaxy mass and in cosmic time than previous studies based on numerical hydrodynamical simulations. Note that only a fraction of the ionizing photons produced within the galaxy actually escape and make their way out into the intergalactic medium. Estimating this 'escape fraction' is extremely difficult, and we do not attempt to address this issue in this chapter.

The roadmap for this chapter is as follows: the semi-analytic framework used in this work is summarized briefly in Section 4.2, where we also describe our procedure for calculating $\dot{N}_{\rm ion}$. We

present our results in Section 4.3, where we compare predictions across several different SPS models (Section 4.3.1), present the scaling relations and redshift evolution for these galaxy populations (Section 4.3.3), and provide specific forecasts for future *JWST* observations in (Section 4.3.3). We then discuss our findings in Section 4.4, and a summary and conclusions follow in Section 4.5.

4.2 The Semi-Analytic Framework

The Santa Cruz semi-analytic model (SAM) for galaxy formation used in this study is very similar to the one outlined in Somerville, Popping, & Trager (2015, hereafter SPT15). The only changes are that we have implemented the updated Okamoto, Gao, & Theuns (2008) photoionization feedback recipe and updated the cosmological parameters to be consistent with the ones reported by the Planck Collaboration in 2015: $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8\,{\rm km\,s^{-1}Mpc^{-1}}$, $\sigma_8 = 0.831$, and $n_s = 0.9665$. We recalibrated the model to $z \sim 0$ observations as described in Chapter 2, where all model components that are essential for this work, along with values of the free parameters, are also documented. We refer the reader to the following works for full details of the modelling framework: Somerville & Primack (1999); Somerville, Primack, & Faber (2001); Somerville et al. (2008, hereafter S08); Somerville et al. (2012, hereafter S12); Popping, Somerville, & Trager (2014, hereafter PST14) and SPT15. Throughout this work, we adopt the cosmological parameters specified above.

The semi-analytic approach to modelling galaxy formation relies on the merger histories of dark matter halos, more commonly known as 'merger trees', which can either be extracted from cosmological N-body simulations or constructed semi-analytically based on the Extended Press-Schechter (EPS) formalism (Press & Schechter 1974; Lacey & Cole 1993). The EPS-based method, which has been shown to be in good qualitative agreement with N-body simulations (Somerville & Kolatt 1999; Somerville et al. 2008; Zhang et al. 2008; Jiang & van den Bosch 2014), is able to efficiently resolve merger histories with very high mass resolution at very low computational cost. We are able to use this method to sample halos over an extremely wide dynamic range, from the ones near the atomic cooling limit to the most massive halos that have collapsed at a given redshift. At each output redshift, we set-up a grid of root halos spanning the range in virial velocity $V_{\rm vir} \approx 20$–$500\ {\rm km\,s^{-1}}$. And for each root halo in the grid, one hundred Monte Carlo realizations of the merger histories are generated, each traced down to progenitors of a minimum resolution mass of either $M_{\rm res} \sim 10^{10}{\rm M}_\odot$ or 1/100th of the root halo mass, whichever is smaller. The expected volume-averaged abundances of these halos are assigned based on the halo mass function (HMF) from the Bolshoi-Planck simulation from the MultiDark suite (Klypin et al. 2016) with fitting functions provided in Rodríguez-Puebla et al. (2016), which has been examined up to $z \approx 10$.

In the latest version of the model (PST14; SPT15), recipes for multiphase gas-partitioning and H_2-based star formation have been implemented and tested. In the former, the disc component of each galaxy is divided into annuli and the cold gas content in each annulus is partitioned into an atomic (HI), ionized (HII), and molecular (H_2) component. In the latter, the surface density of SFR ($\Sigma_{\rm SFR}$) is estimated based on the surface density of molecular hydrogen ($\Sigma_{\rm H_2}$) using observationally motivated, empirical H_2-based SF relations. SPT15 investigated several different gas partitioning and SF recipes, and favoured the metallicity-based, UV-background-dependent gas partitioning recipe, which is based on simulations by Gnedin & Kravtsov (2011, hereafter GK). In the star formation recipe, $\Sigma_{\rm SFR}$ is expressed as a broken power law function of $\Sigma_{\rm H_2}$. SPT15 investigated a power law with a single slope, and one in which the slope becomes steeper above a critical H_2 surface density ('two-slope'). They found that both recipes produce similar results at $z \sim 0$, but at high redshift ($z \sim 4$–6), the 'two-slope' recipe was favoured, in agreement with direct evidence from recent observations (e.g. Sharon et al. 2013; Rawle et al. 2014; Hodge et al. 2015; Tacconi et al. 2018). In Chapter 2 and Chapter 3, we investigated both star formation laws up to $z \sim 10$, and concluded even more strongly that the 'two-slope' relation is much more consistent with observations of luminous high-redshift galaxies. In this chapter and the remainder of the thesis, we therefore adopt the 'two-slope' star formation relation, which we have referred to as the 'fiducial' GK-Big2 model throughout this thesis.

Our model consists of a collection of physically motivated, phenomenological and empirical recipes, including cosmological accretion and cooling, stellar-driven winds, chemical evolution, black hole growth and feedback, galaxy mergers, etc. These recipes are then used to track the evolution of a wide range of global physical properties of galaxies, such as stellar mass, star formation rate, masses of multiple species of gas (ionized, atomic, molecular), and stellar and gas phase metallicity, etc. For each galaxy, SEDs are created by convolving the predicted stellar age and metallicity with SPS models of Bruzual & Charlot (2003, hereafter BC03) with the Padova1994 (Bertelli et al. 1994) isochrones assuming a universal Chabrier stellar IMF (Chabrier 2003a). Throughout this work, all magnitudes are expressed in the AB system (Oke & Gunn 1983) and all uses of log are base 10 unless otherwise specified.

4.2.1 Calculating $\dot{N}_{\rm ion}$

As a useful reference, in Fig. 4.1 we first show the dependence of $\dot{N}_{\rm ion}$, $L_{\rm UV}$, and $\xi_{\rm ion}$ as a function of age for simple stellar populations, assuming an instantaneous starburst, of different metallicities ($Z = 0.0001$, 0.004, 0.008, and 0.020), in two popular sets of SPS models, BC03 and BPASS. We show the BPASS models that account only for single star evolution, as is also the case in the BC03 models,

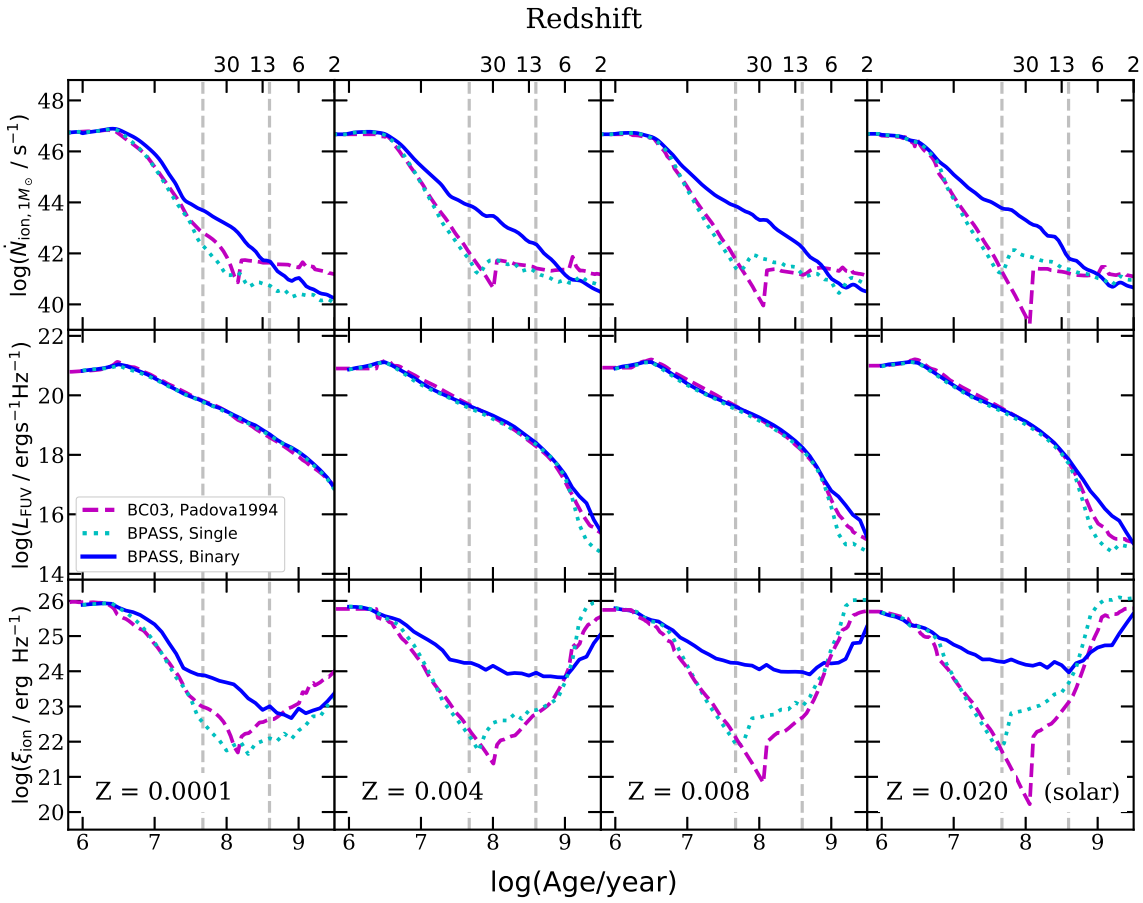

Figure 4.1 The evolution tracks of the production rate of ionizing photons per solar mass of stars formed ($\dot{N}_{\rm ion,1M_\odot}$; top row), the far-UV luminosity $L_{\rm FUV}$ (also normalized to $1{\rm M}_\odot$; middle row), and the production efficiency of ionizing photons ($\xi_{\rm ion}$; bottom row) as a function of the age of a stellar population, assuming an instantaneous starburst, predicted by the BC03 (purple dashed) and BPASS single (cyan dotted) and binary (blue solid) SPS models. Each column shows a different metallicity as labelled. We also show the redshifts corresponding to the age equal to the age of the Universe, which can be taken as the upper limit on the stellar population age at that redshift. The vertical grey dashed lines show the age range that is most relevant to this work, where the upper (lower) bound is marked by the median of the mass-weighted stellar age for $z = 4$ ($z = 10$) galaxy populations predicted by our model.

and those that account for binary star evolution. These quantities are directly obtained from the data tables released by the BC03[1] and BPASS groups[2] (Stanway et al. 2016; Eldridge et al. 2017; Stanway & Eldridge 2018). Both SPS models adopted in this work assume a Chabrier IMF with an upper mass cut-off $m_{\rm U} = 100{\rm M}_\odot$. In the top panel of Fig. 4.1, we show the evolution of ionizing photon production rate normalized to that for a single solar mass of stars, $\dot{N}_{\rm ion,1M_\odot}$. As discussed by Stanway et al. (2016), this quantity is strongly affected by both binary evolution and metallicity. However, in the middle panel of Fig. 4.1 we see that differences in rest-frame far-ultraviolet (FUV) luminosity, $L_{\rm FUV}$, predicted by different models are rather small for young stellar populations. As a result, in the bottom panel, which shows $\xi_{\rm ion}$, we can see that the value of this quantity can differ by as much as several orders of magnitude at intermediate ages for models that include binary evolution relative to those that do not.

In this work, we convolve these SPS predictions for $\dot{N}_{\rm ion}$ with the predicted 2D histogram of stellar ages and metallicities present in each galaxy at a given output redshift. These histograms are built up as follows. At each time-step during each galaxy's evolution, a 'star parcel' dm_* is created with a mass determined by the H_2 density in the disc and the H_2-based SF recipe. Each star parcel inherits the metallicity of the cold gas in the ISM at the time it forms. Metals are deposited in the cold gas by new stars with an assumed yield, and may be removed from the ISM by stellar driven winds (see SPT15 for details). When galaxies merge, their stellar populations are combined. At an output redshift, we evaluate $\dot{N}_{\rm ion}$ for each galaxy by summing up $\dot{N}_{\rm ion}$ for each star parcel from the data tables provided by a chosen SPS model based on its stellar age and metallicity at the time.

In the previous chapters in this thesis, we adopted the BC03 SPS models to compute the galaxy SEDs, and the model UV luminosity functions (UV LFs) have been extensively tested against existing observations. In order to avoid repeating these comparisons, we retain the estimates of $L_{\rm FUV}$ based on the BC03 models in all of our results, and only compute $\dot{N}_{\rm ion}$ with the SPS model variants. Thus the values of $\xi_{\rm ion}$ for the BPASS models are not strictly self-consistent. However, we saw in Fig. 4.1 that the predictions for $L_{\rm FUV}$ from these two SPS models are very similar for the ages that are most relevant for our study ($z \gtrsim 6$), so this should not cause a significant discrepancy. Furthermore, $\xi_{\rm ion}$ is only used for illustrative purposes. For actually computing the reionization history, we use $\dot{N}_{\rm ion}$ directly.

[1] http://www.bruzual.org/ gbruzual/bc03/
[2] https://bpass.auckland.ac.nz/, v2.2.1

4.3 Results

Our results are organized in three parts. 1) We compare results across different SPS models. 2) We present scaling relations and redshift evolution for ionizing photon production rates and efficiencies, and show how $\xi_{\rm ion}$ relates to observable properties. 3) We show how variations in the SN feedback model impact our predictions for galaxies during the EoR.

4.3.1 Comparison across SPS models

In this subsection, we compare $\xi_{\rm ion}$ computed with several SPS models, while other model components in our SAM remain in their fiducial configurations. As illustrated in Fig. 4.1, the predicted production rate of ionizing photons depends on the age and metallicity of the stellar populations, and these relations vary across different SPS models. Note that these SPS models are applied to the same galaxy populations. In other words, the galaxies used for all three SPS models have identical star formation histories and physical properties.

In Fig. 4.2, we show the distributions of $\xi_{\rm ion}$ against M_* for BC03 and BPASS models for single and binary stellar population at selected redshifts. The 2D histograms illustrate the conditional number density per Mpc^3 of galaxies in each bin and are normalized to the sum of the number density in the corresponding (vertical) stellar mass bin. The 16th, 50th, and 84th percentiles are marked in each panel to illustrate the statistical distribution. The median of predictions of all models for each given redshift are overlaid for easy comparison.

In general, we see that massive galaxies have lower $\xi_{\rm ion}$ than low-mass galaxies regardless of the choice of SPS models and redshift. Also the scatter of the relation is larger for low-mass galaxies. Although all models demonstrate some level of dependence on M_*, we find that BC03 yields the weakest dependence for $M_* < 10^8$ M$_\odot$, but both BPASS models predict more evolution for galaxies in the same mass range. The differences between predictions from the most and the least optimistic SPS model can be up to ~ 0.2 dex for galaxies with $M_* \sim 10^7$M$_\odot$, and the difference shrinks for more massive galaxies.

At fixed M_*, the predicted value of $\xi_{\rm ion}$ evolves mildly as a function of redshift, with an average downward shift of ~ 0.1 dex from $z = 10$ to $z = 4$. High-redshift galaxies seem to be more efficient at producing ionizing photons than their low-redshift counterparts of similar mass due to the younger stellar age and lower metallicity. We also notice the slope of the $\xi_{\rm ion}$-M_* relation becomes more shallow with decreasing redshift, likely due to the corresponding change in the slope of the M_*-Z_* relation.

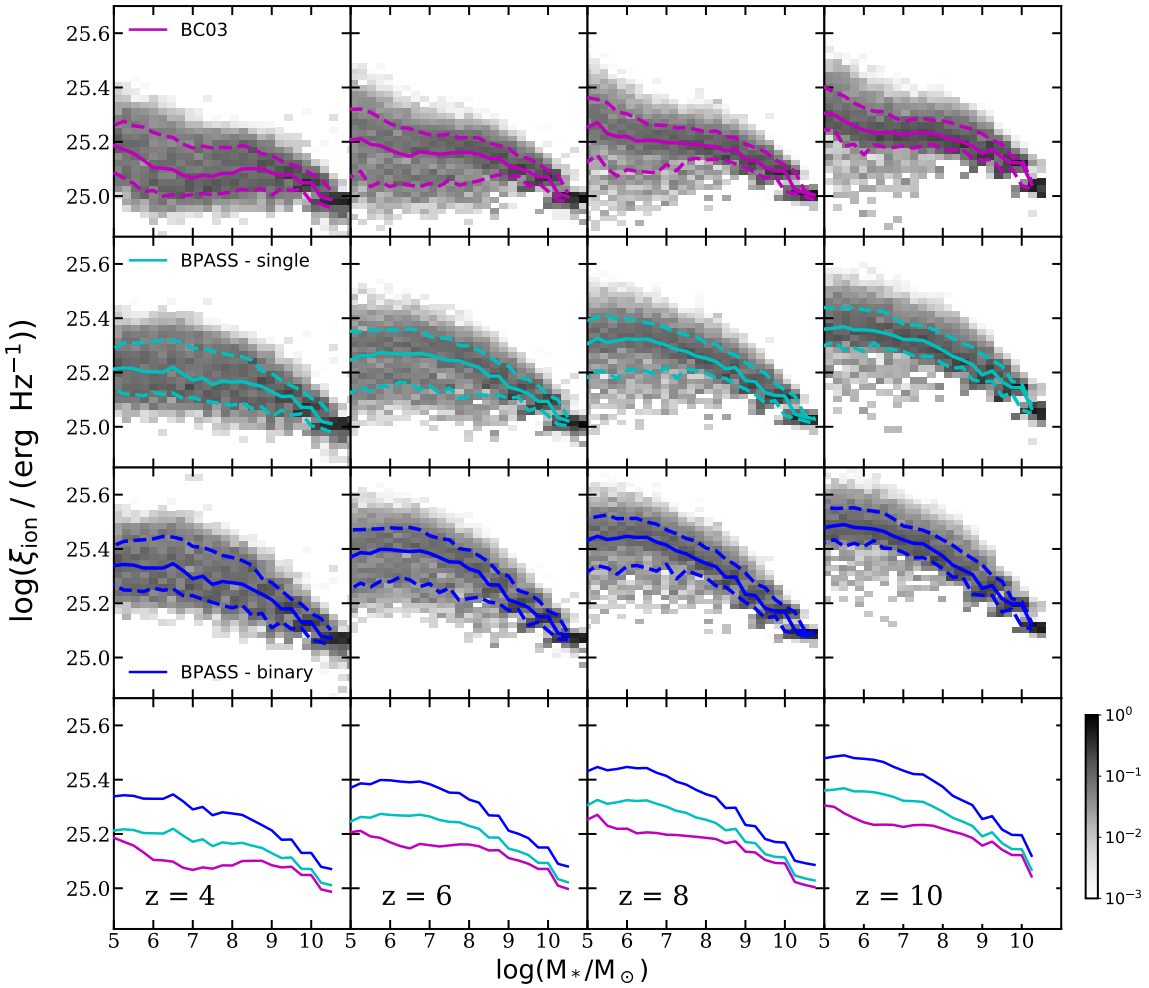

Figure 4.2 Conditional distributions of $\xi_{\rm ion}$ versus stellar mass predicted at $z = 4$, 6, 8, and 10 using SPS models from BC03 (purple) and BPASS (cyan and blue for single and binary). Each column shows a different redshift as labelled. The grey-scale 2D histograms show the conditional number density of galaxies in each bin, normalized to the sum of the number density per Mpc^3 in each corresponding (vertical) stellar mass bin.The blue solid and dashed lines show the 50th, 16th and 84th percentiles, respectively, for the distribution in each panel. The medians of these distributions are repeated in the bottom row for comparison, using the same colour coding as in the other panels.

4.3.2 Comparison with numerical hydrodynamic simulations

We compare our results with predictions of $\xi_{\rm ion}$ computed from galaxies in the BlueTides simulations with various SPS models presented in W16. BlueTides is a cosmological hydrodynamical simulation with 400 Mpc h^{-1} on a side and 2×7040^3 particles, which is able to resolve galaxies with $M_* \gtrsim 10^8$ providing a wide variety of predictions for physical and observable properties for high-redshift galaxies up to $z \sim 8$ (Feng et al. 2016; Wilkins et al. 2017). We also note that BlueTides adopted a set of cosmological parameters that are consistent with constraints from WMAP9. In Fig. 4.3, we compare the $\xi_{\rm ion}$ relations at $z = 8$ predicted by our SAM and BlueTides with BC03 and bpass single and binary star models (v2.0, Stanway et al. 2016), which does not include nebular emission. The W16 calculations assumed a Salpeter (1955) IMF. We convert the quoted stellar masses to a Chabrier IMF by adding -0.21 dex for direct comparison (Madau & Dickinson 2014). The values of $\xi_{\rm ion}$ are nearly unaffected by this change in IMF. We note that the predictions for $\xi_{\rm ion}$ from W16 and our models are very consistent for the BC03 models, while the SAM predictions are about 0.1 dex lower for the bpass single and about 0.25 dex lower for the bpass binary models (though the qualitative trends are the same).

In order to better understand these differences, we further compare a number of key physical properties for our model galaxies to the predictions by BlueTides reported in Wilkins et al. (2017) in Fig. 4.4, including intrinsic (dust-free) UV mass-to-light ratio, stellar metallicity, mass-weighted stellar age, and specific star formation rate (sSFR). Note that this work assumed a Chabrier IMF, which is consistent with our work and does not require any adjustment. We also compare to the mass-metallicity relation predicted by the FIRE (Feedback in Realistic Environment) simulations (Hopkins et al. 2014; Ma et al. 2016a), which are high-resolution cosmological hydrodynamic zoom-in simulations. We show predictions at $z = 6$ in our comparison, which is the highest redshift provided in the FIRE publications. The FIRE simulations assumed a Kroupa (2001) IMF and we convert their stellar masses to a Chabrier IMF by adding -0.03 dex Madau & Dickinson (2014). It is interesting that although BlueTides and FIRE have very different resolution as well as different subgrid recipes for star formation and stellar feedback, the predicted mass-metallicity relations are very similar, and much lower than the one predicted by our SAM. We discuss possible reasons for this discrepancy in Section 4.4, however, we do not currently have a complete understanding of this issue, and it is beyond the scope of this thesis to determine.

The differences in the stellar population properties between the SAM and BlueTides galaxies do not seem sufficient to explain the difference in the predicted value of $\xi_{\rm ion}$, nor it is obvious how this would explain why the differences are more pronounced for the bpass models. However, W16

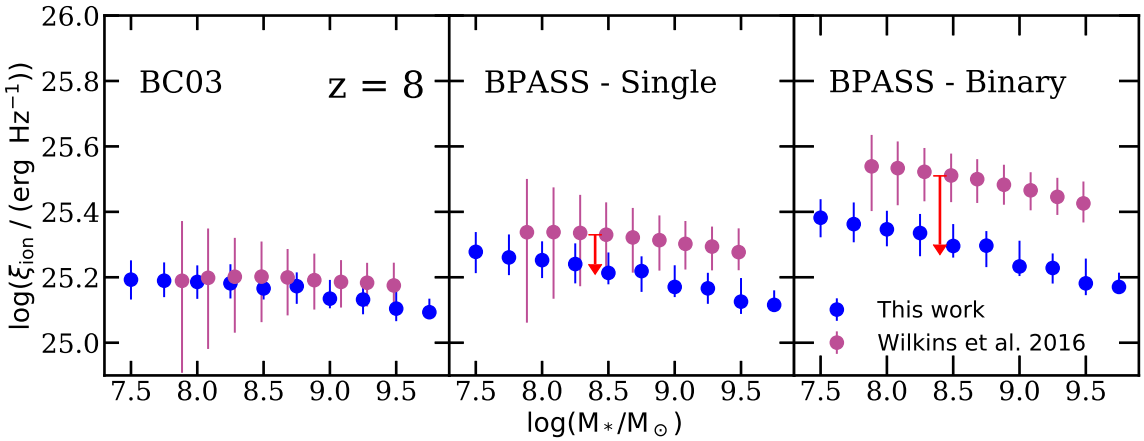

Figure 4.3 A comparison of $\xi_{\rm ion}$ predicted with BC03 and BPASS (both single and binary) SPS models between this work and W16. The data points mark the median and the error bars span the 16th to 84th percentile range. The red arrow shows the estimated difference in $\xi_{\rm ion}$ changing from BPASS v2.0 to v2.1 models, accounting for the metallicities difference between the SAM and BlueTides galaxies, see text for details.

used the BPASS v2.0 version of the models, while we use the latest v2.2.1 version. Eldridge et al. (2017) provide a detailed discussion of the implications of changes from the v2.0 to v2.1 models for the production efficiency of ionizing photons (see their section 6.6.1). They find that the use of improved stellar atmosphere models for low-metallicity O stars leads to a less dramatic increase in $\xi_{\rm ion}$ from the single star to binary star case than was seen in the v2.0 models, as well as a weaker dependence of $\xi_{\rm ion}$ on metallicity. We make use of their figure 35 to estimate the effect of changing from the v2.0 to v2.1 models, assuming the typical metallicity of the SAM and BlueTides galaxies. This correction is shown by the red arrow in our Fig. 4.3, showing that this largely accounts for the discrepancy.

4.3.3 Scaling relations for the production rate and efficiency of ionizing radiation

In this subsection, we present scaling relations for $\xi_{\rm ion}$ and $\dot{N}_{\rm ion}$ with respect to selected observable and physical properties predicted by our fiducial model configurations with the BPASS binary SPS models. We also study the section: these relations and compare to recent observations.

We show the distribution of $\xi_{\rm ion}$ against SFR (averaged over 100 Myr), specific SFR (sSFR), intrinsic rest-frame $M_{\rm UV}$ (without attenuation by dust), and cold gas metallicity ($Z_{\rm cold}$) predicted at $z = 4$, 6, 8, and 10 in Fig. 4.5. Throughout this work, we refer to the UV luminosity without a correction for attenuation effect by dust as the intrinsic $M_{\rm UV}$, which is to distinguish it from the dust-attenuated $M_{\rm UV}$. Similar to plots in previous sections, the 2D histograms indicate the conditional number density per Mpc^3 of galaxies in each bin and is normalized to the sum of the

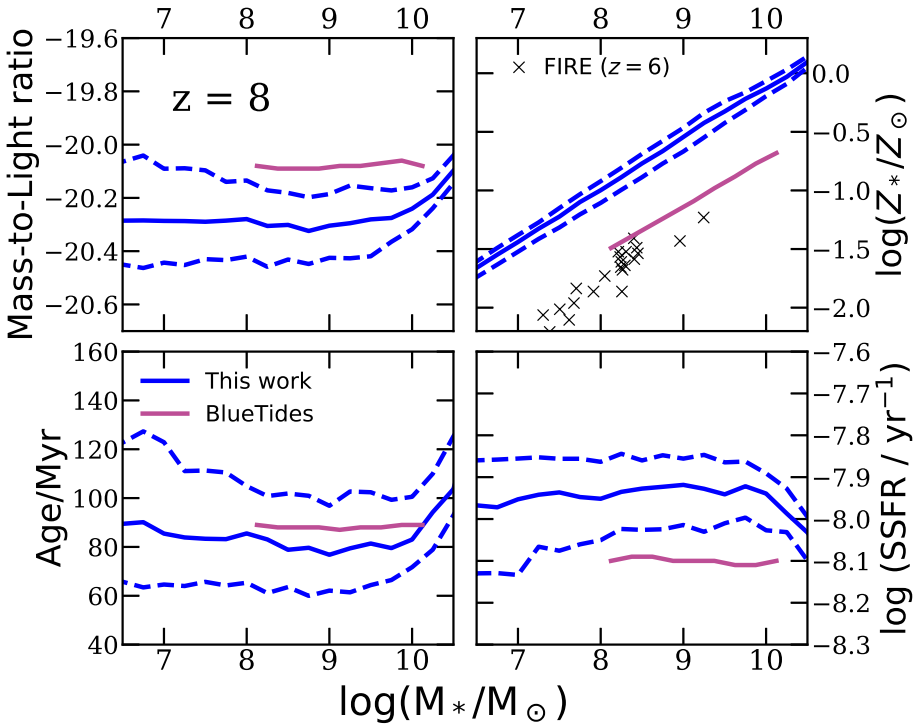

Figure 4.4 Comparison between the median of the intrinsic rest-UV mass-to-light ratio ($\log((M_*/\mathrm{M}_\odot)/(L_{\nu,\mathrm{FUV}}/\mathrm{erg\ s^{-1}Hz^{-1}}))$; upper left), stellar metallicity (upper right), mass-weighted stellar age (lower left), and specific SFR (lower right) for galaxies predicted by our SAM (blue) and BlueTides (Wilkins et al. 2017, purple). The black crosses show the stellar metallicity predicted by the FIRE simulations at $z = 6$ (Ma et al. 2016a). The dashed blue lines mark the 16th and 84th percentiles of the SAM predicted distributions.

number density in its corresponding (vertical) stellar mass bin. The 16th, 50th, and 84th percentiles are marked in each panel. The median of predictions made with all models for each given redshift are overlaid to illustrate the redshift evolution.

We find mild trends of decreasing $\xi_{\rm ion}$ with increasing SFR, increasing $\xi_{\rm ion}$ with increasing sSFR, decreasing $\xi_{\rm ion}$ with increasing UV luminosity, and decreasing $\xi_{\rm ion}$ with increasing gas-phase metallicity. The inverse trends of $\xi_{\rm ion}$ with SFR and UV luminosity may appear counterintuitive, until we recall that both of these quantities are strongly correlated with stellar mass, and hence with metallicity, so the inverse correlation between metallicity and $\xi_{\rm ion}$ is likely driving this result. As seen in Fig. 4.4, there is only a weak trend between stellar mass and mean stellar age in our models. At fixed SFR, UV luminosity, and cold gas metallicity, $\xi_{\rm ion}$ decreases with cosmic time from $z \sim 10$ to 4, due to the increasing fraction of relatively old stars. Only for fixed sSFR is $\xi_{\rm ion}$ nearly constant with redshift, indicating that this quantity is correlated with the fraction of young stars.

Fig. 4.6 shows the distribution of $\xi_{\rm ion}$, predicted for halo populations between $z = 4$–10. We see that for halos of the same mass, the production efficiency of ionizing photons can change by up to ~ 0.2 dex between $z = 10$ and 4. Similarly, Fig. 4.7 shows the distribution of the specific ionizing photon production rate, $\dot{N}_{\rm ion}/M_{\rm h}$, predicted for the same set of halos, where the production rate of ionizing photons can be up to ~ 1.5 orders of magnitude higher at $z = 10$ relative to $z = 4$ for halos of the same mass. This relation evolves much more rapidly than the stellar-to-halo mass ratio (SHMR) presented in Chapter 3. This is due to our finding (presented in Chapter 3) that galaxies at high redshift are intrinsically brighter in the UV than their low-redshift counterparts of similar stellar mass due to the higher SFR, younger stellar population, and lower metallicities.

On the other hand, $\xi_{\rm ion}$ does not evolve as much across redshift since both $\dot{N}_{\rm ion}$ and $L_{\rm FUV}$ evolve in the same direction. In Fig. 4.8, we illustrate the evolution of $\xi_{\rm ion}$ as a function of redshift. We show predictions from our fiducial model with both BPASS and BC03 SPS models at the bracketing UV magnitudes of $M_{\rm UV} = -22$ and -17 and compare these to a compilation of observational estimates. Both our predictions and the observations are colour-coded for rest-frame $M_{\rm UV}$. These measurements are derived from Hα flux and UV-continuum luminosity (Bouwens et al. 2016c; Smit et al. 2016; Rasappu et al. 2016; Lam et al. 2019), Hα flux and UV-based SFR (Shim et al. 2011; Mármol-Queraltó et al. 2016), CIV measurement (Stark et al. 2015), and SED fitting (Stark et al. 2017). We indicate the range of $\log \xi_{\rm ion} = 24.0$ to 25.7 derived from ∼14,000 compact star-forming galaxies found in the Sloan Digital Sky Survey (SDSS; DR12) between $0 < z < 1$ for comparison (Izotov et al. 2017). We also include the luminosity-weighted $\xi_{\rm ion}$ from W16 at = 8–10. We compare our predictions to the results from the FirstLight simulations, which is a set of high-resolution zoom-in simulations based on cosmological boxes of 10, 20, and 40 Mpc h^{-1} on a side, resolving galaxies

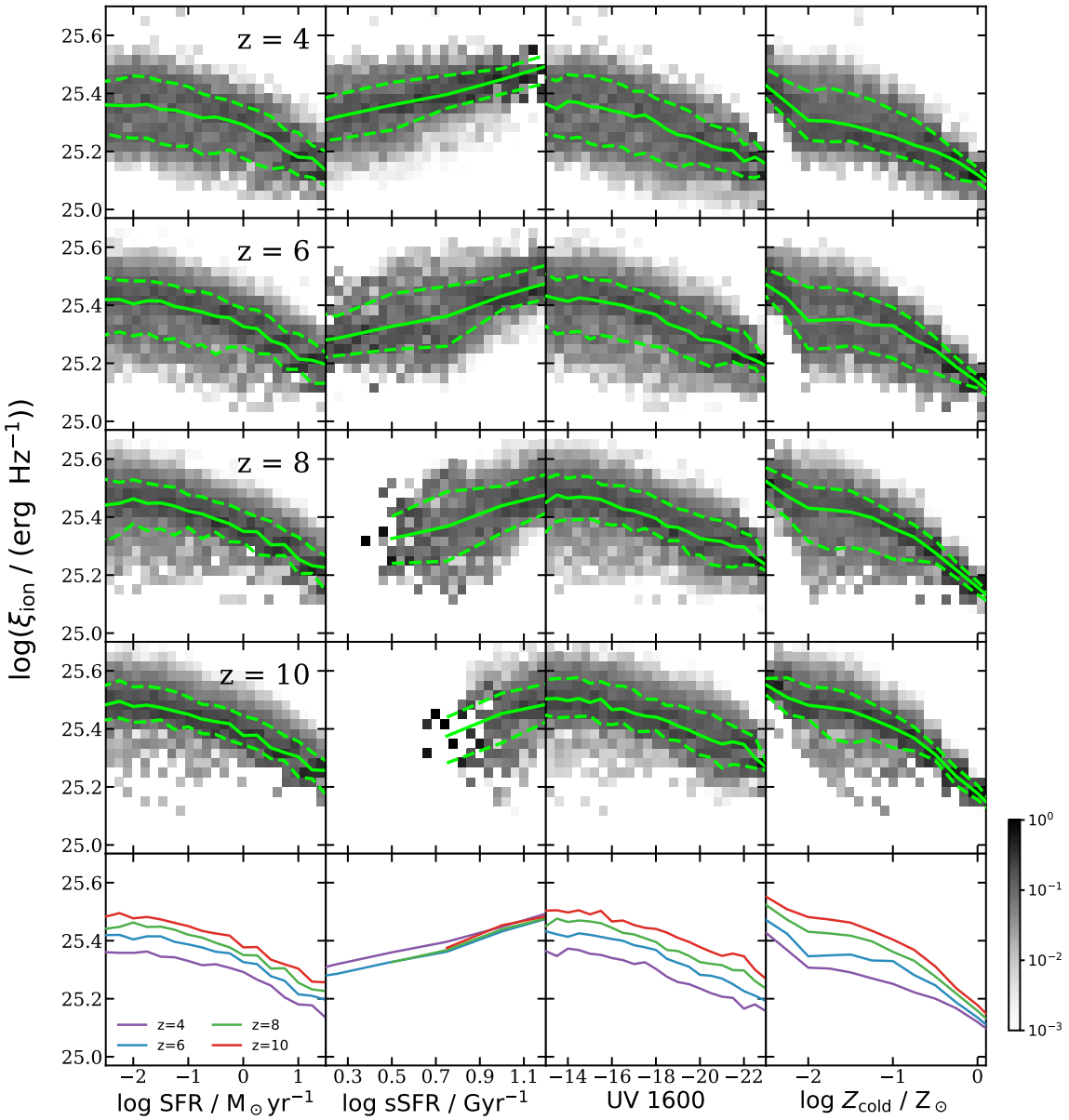

Figure 4.5 Conditional distributions of $\xi_{\rm ion}$ versus SFR (averaged over 100 Myr), specific SFR (sSFR), intrinsic UV luminosity, and cold gas metallicity at $z = 4$, 6, 8, and 10 with our fiducial model configurations with the BPASS binary SPS models. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The greyscale 2D histograms show the conditional number density of galaxies in each bin, normalized to the sum of the number density per Mpc^3 in the corresponding (vertical) bin. The median of these distributions are repeated in the last row for comparison to illustrate the redshift evolution.

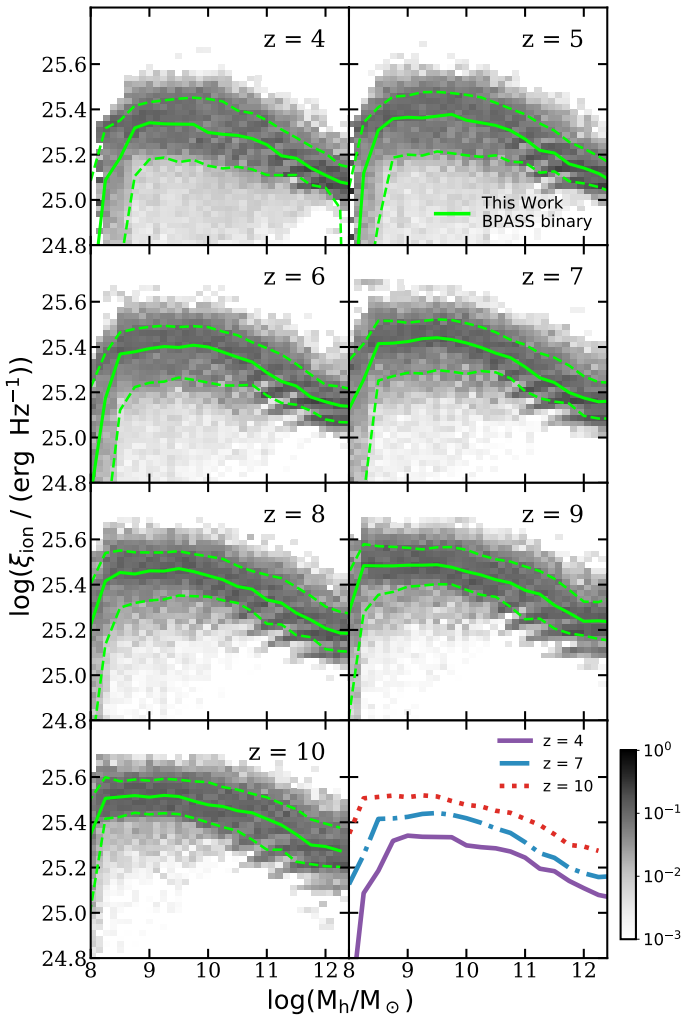

Figure 4.6 Predicted $\xi_{\rm ion}$ as a function of halo mass between $z = 4$–10 predicted with the BPASS binary model. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The 2D histograms are colour-coded to show the conditional number density per Mpc^3 of galaxies in each bin, normalized to the sum of the number density in the corresponding (vertical) halo mass bin. The last panel show an overlay of the SHMR median predicted at $z = 4$, 7, and 10 from our model.

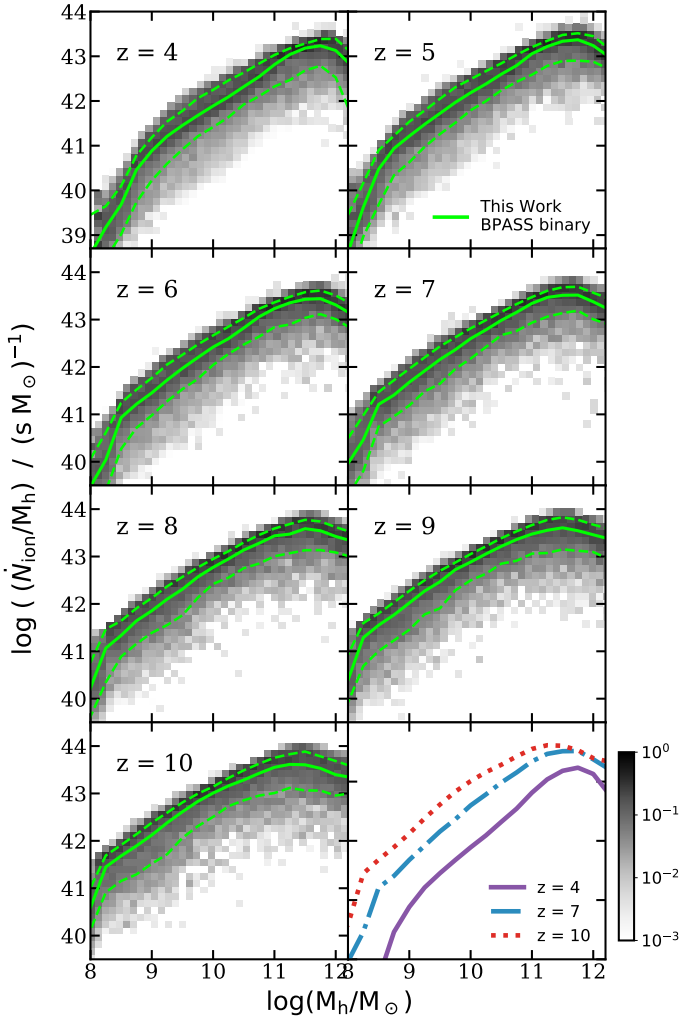

Figure 4.7 Specific ionizing photon production rate, $\dot{N}_{\rm ion}/M_{\rm h}$, as a function of halo mass between $z = 4$–10 predicted with the BPASS binary model. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The greyscale 2D histograms show the conditional number density per Mpc^3 of galaxies in each bin, normalized to the sum of the number density in the corresponding (vertical) halo mass bin. The last panel shows the median $\dot{N}_{\rm ion}/M_{\rm h}$ relations predicted at $z = 4$, 7, and 10 from our model.

in halos with circular velocities spanning the range between 50–250 km s^{-1} at $z \gtrsim 5$ (Ceverino et al. 2017, 2019). The computation for $\xi_{\rm ion}$ is based on synthetic SEDs from BPASS (v2.1, Eldridge et al. 2017) including nebular emission (Xiao, Stanway, & Eldridge 2018) and a Kroupa-like IMF. Note that whilst the simulations have been run with both WMAP and Planck cosmologies, the results referenced here are only available with the WMAP-5 cosmology. For reference, we also label the range of typical values $\xi_{\rm ion} \approx 25.20$–$25.30$ that are assumed in many analytic studies (e.g. Bouwens et al. 2012; Finkelstein et al. 2012a; Kuhlen & Faucher-Giguère 2012; Duncan & Conselice 2015; Robertson et al. 2015). Our models predict very mild evolution in $\xi_{\rm ion}$ as a function of redshift and a very consistent ~ 0.1–0.15 dex difference between the bright ($M_{\rm UV} = -22$) and faint ($M_{\rm UV} = -17$) galaxies. The red arrow shows an estimate (from figure 35 of Eldridge et al. (2017)) of the increase in $\xi_{\rm ion}$ if the upper mass limit of the IMF were changed from 100 M$_\odot$ to 300 M$_\odot$, illustrating how sensitive $\xi_{\rm ion}$ is to the upper end of the stellar IMF. In addition, lower redshift measurements reporting $\xi_{\rm ion} \gtrsim 25.5$ at $z \sim 3$ appear to be in tension with standard model assumptions (Nakajima et al. 2016, 2018). We also note that estimates of $\xi_{\rm ion}$ from detected rest-UV emission lines at higher redshifts (e.g. $z > 5$) tend to be obtained for sources with detectable nebular lines. As there is a general correlation between the equivalent width (EW) in these lines and $\xi_{\rm ion}$, the published $\xi_{\rm ion}$ values based on rest-UV lines are likely biased towards the upper end of the true range of values, rather than reflecting the mean or median of the overall population at these epochs.

F19 obtained indirect constraints on the plausible range of $\xi_{\rm ion}$ using an empirical model. They parametrization $\xi_{\rm ion}$ as a function of redshift and rest-frame $M_{\rm UV}$, and obtained the posterior on these parameters (along with other model parameters) using Markov Chain Monte Carlo (MCMC). They found that it was possible to satisfy constraints on the reionization history of the Universe with a relatively low average escape fraction (of about 5 percent) if $\xi_{\rm ion}$ is higher in lower luminosity and higher redshift galaxies. The median posterior parameters found in their analysis are $d\log\xi_{\rm ion}/dz = 0.13$ and $d\log\xi_{\rm ion}/dz = 0.08$, and a reference magnitude of $M_{\rm UV,ref} = 20$. Although our models predict these qualitative trends, our predicted trends are not nearly as strong as the ones suggested by the F19 models. We will investigate resolutions to this tension in more detail in Chapter 5 of this thesis.

Next we attempt to make predictions for correlations between physical and observable galaxy properties and $\xi_{\rm ion}$ in order to suggest strategies for future observations. In Fig. 4.9, we show observed-frame magnitude in the NIRCam F200W filter, $m_{\rm F200W}$, vs. the UV-continuum slope, $\beta_{\rm UV}$, in the xy-plane and colour code each cell by $\xi_{\rm ion}$. The detection limits for *JWST* wide- and deep-field observations are also marked for quick reference, which are obtained assuming survey configurations that are comparable to those of past *HST* surveys and the published NIRCam sensitivities with

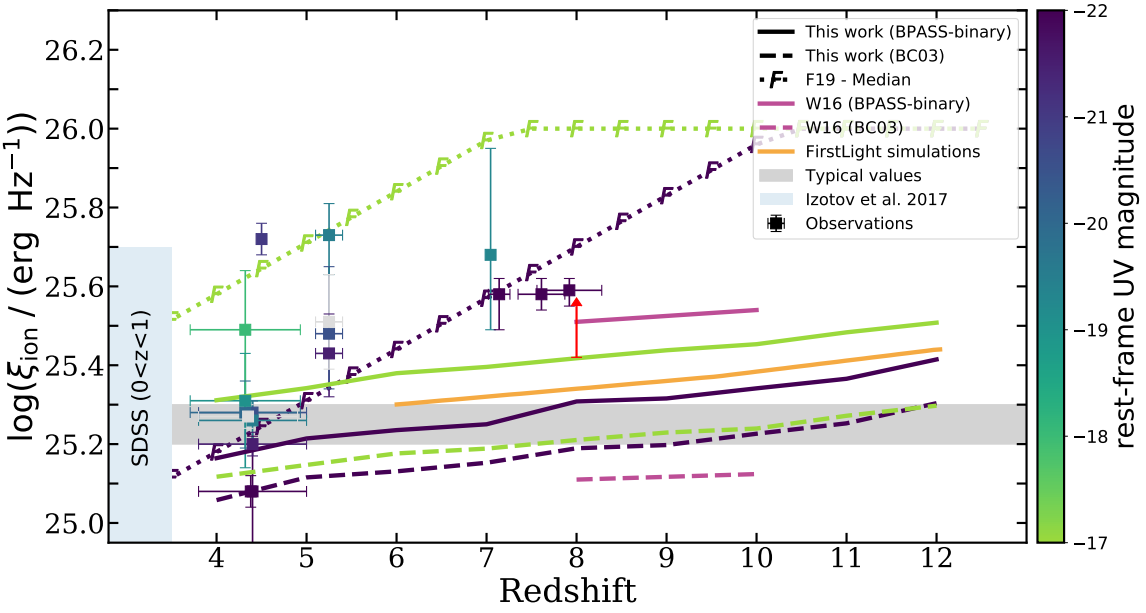

Figure 4.8 A comparison of the redshift evolution of $\xi_{\rm ion}$ predicted by our models to a compilation of observations and empirical constraints. The solid and dashed lines denote our predictions using BC03 (solid) and the BPASS binary SPS (dashed) models and are colour-coded for rest-frame dust-attenuated $M_{\rm UV}$. Here we show the two bracketing cases representing bright ($M_{\rm UV} = -22$) and faint ($M_{\rm UV} \sim= -17$) galaxies shown in dark blue and green, respectively. The red arrow shows an estimate of the increase in $\xi_{\rm ion}$ if the upper mass limit of the IMF were changed from 100 $M_\odot$ to 300 $M_\odot$. We also plot the median parameters obtained by F19 for the same bracketing $M_{\rm UV}$, shown with a dotted line with a 'F' marker. The solid and dashed purple lines show the luminosity-weighted value of $\xi_{\rm ion}$ for galaxies with $M_* > 10^8$ $M_\odot$ at $z = 8$–10 predicted by W16 with the SPS models considered in this work. The solid orange line shows results from the FirstLight simulations, which assumed a BPASS binary SPS model (Ceverino et al. 2019). We include a compilation of observations from Shim et al. (2011); Stark et al. (2015, 2017); Bouwens et al. (2016c); Smit et al. (2016); Rasappu et al. (2016); Lam et al. (2019) for comparison; see text for a description. All observations and model predictions are colour-coded for rest-frame UV magnitude. Typical values adopted in analytic calculations are highlighted by the grey band. We also show the range derived from a large sample of SDSS galaxies at $0 < z < 1$ (Izotov et al. 2017). The lower bound of this range extends to $\xi_{\rm ion} = 24$ (not shown).

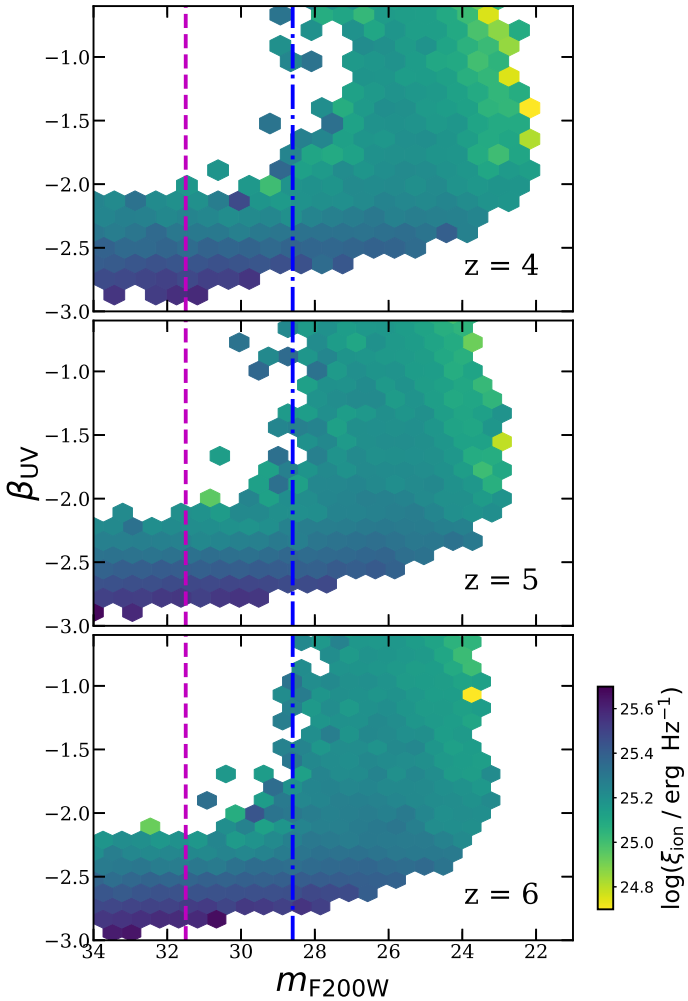

Figure 4.9 Correlation of $\beta_{\rm UV}$ and $m_{\rm F200W}$ at $z = 4$, 5, and 6 in our fiducial models, with the average value of $\xi_{\rm ion}$ in each hexbin colour-coded. The vertical lines mark the detection limit of a typical *JWST* NIRCam wide and deep surveys at $m_{\rm F200W,\ lim} = 28.6$ and 31.5, respectively. The left edges of these plots coincide with the detection limit for lensed surveys.

the F200W filter (see Chapter 2). We present these predictions for $z = 4$, 5, and 6, where many galaxies should be bright enough that spectra will be relatively efficiently obtained with *JWST* or other facilities. We show a similar plot for sSFR and stellar metallicity in Fig. 4.10.

4.3.4 The impact of SN feedback

As shown in previous chapters in this thesis, the number density of faint, low-mass galaxies is very sensitive to the feedback strength, as stronger feedback suppresses star formation by ejecting gas from the ISM. However, the faint- and low-mass-end slope of UV LFs and SMF at high redshift are not well constrained by current observations. In Chapter 2 and Chapter 3, we experimented with alternative SN feedback slopes $\alpha_{\rm rh}$, which effectively characterize the dependence of mass outflow

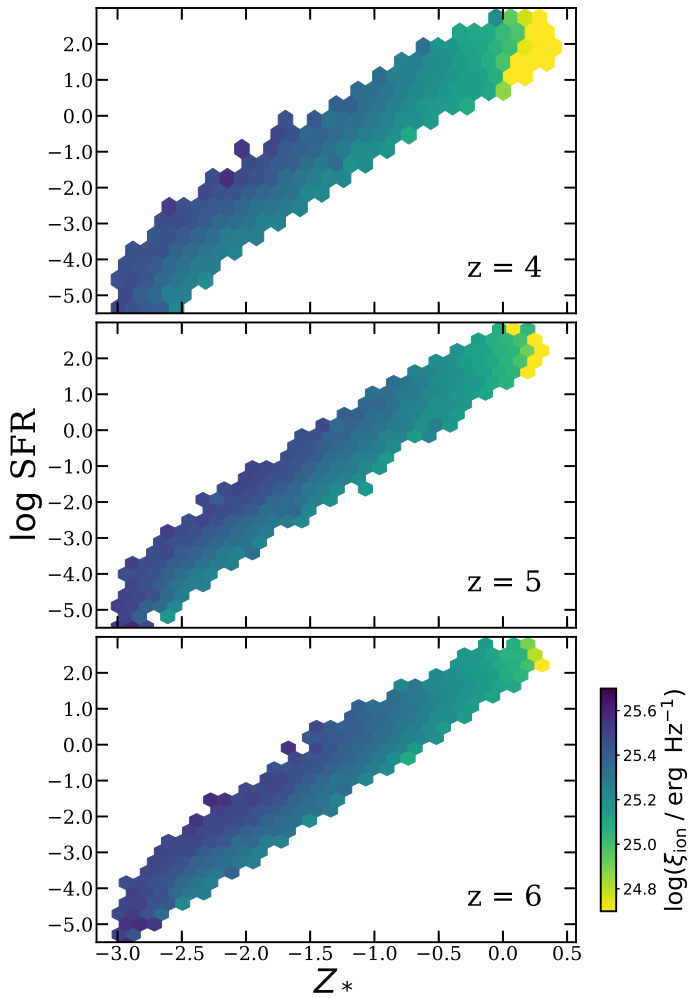

Figure 4.10 Correlation of stellar metallicity and SFR at $z = 4$, 5, and 6 in our fiducial models, with the average value of $\xi_{\rm ion}$ in each hexbin colour-coded.

rate on circular velocity, and found that a fiducial value of $\alpha_{\rm rh} = 2.8$ yielded predictions that best match existing observational constraints. We further found that a range of values from $\alpha_{\rm rh} = 2.0$ (weaker feedback) to $\alpha_{\rm rh} = 3.6$ (stronger feedback) yielded predictions that are well within the current observational uncertainties. In this subsection, we explore how varying this model parameter impacts the production of ionizing photons.

Fig. 4.11 presents the fractional difference of the median of the predicted $\dot{N}_{\rm ion}$ between our fiducial model ($\alpha_{\rm rh} = 2.8$) and model variances with stronger ($\alpha_{\rm rh} = 3.6$) and weaker ($\alpha_{\rm rh} = 2.0$) SN feedback, computed using a sliding boxcar filter of width $\Delta \log(M_*/{\rm M}_\odot) = 1.0$ in stellar mass. We find that the range of $\alpha_{\rm rh}$ that yields galaxy populations within the observational uncertainties may cause $\sim 20\%$ differences in $\dot{N}_{\rm ion}$, where more ionizing photons are produced when feedback is stronger and vice versa. This is likely caused by the more bursty star formation in the presence of stronger feedback. Although stronger feedback may boost the production *efficiency* of ionizing photons for low-mass galaxies, we must keep in mind that stronger feedback also suppresses the formation of low-mass galaxies, leading to lower overall UV luminosity density, so there is a trade-off. We show in Fig. 4.12 and in Chapter 5 that in our models, stronger feedback leads to an overall decrease in the total number of ionizing photons. However, different implementations of stellar feedback could in principle produce different results.

4.4 Discussion

In this work, we compute the production rate $\dot{N}_{\rm ion}$ and production efficiency $\xi_{\rm ion}$ of ionizing photons by combining stellar population synthesis models with a physically grounded model of galaxy formation set in a cosmological framework, the Santa Cruz SAM. Our fiducial model incorporates multiphase gas partitioning and H_2-based star formation recipes (GK-Big2), which have been shown to reproduce observational constraints over a wide range of redshifts from $z = 0$–10 (see PST14, SPT15, Chapter 2, and Chapter 3). $\dot{N}_{\rm ion}$ and $\xi_{\rm ion}$ for composite stellar populations depends on the joint distribution of ages and metallicities in each galaxy. These quantities may in principle evolve with redshift or be correlated with global galaxy properties such as mass or luminosity. It is thus important to use a self-consistent, physically grounded model that hopefully contains the relevant internal correlations between physical properties. This is a significant advantage of our approach.

We compare $\dot{N}_{\rm ion}$ and $\xi_{\rm ion}$ calculated using different SPS models, which depends on the physical properties of the stellar populations, such as star formation and metal enrichment histories. This approach of modelling $\dot{N}_{\rm ion}$ also makes physical predictions for the correlations among $\xi_{\rm ion}$ and $\dot{N}_{\rm ion}$ with other physical properties, such as M_*, Z_*, SFR, etc., and observable properties, such as $\beta_{\rm UV}$

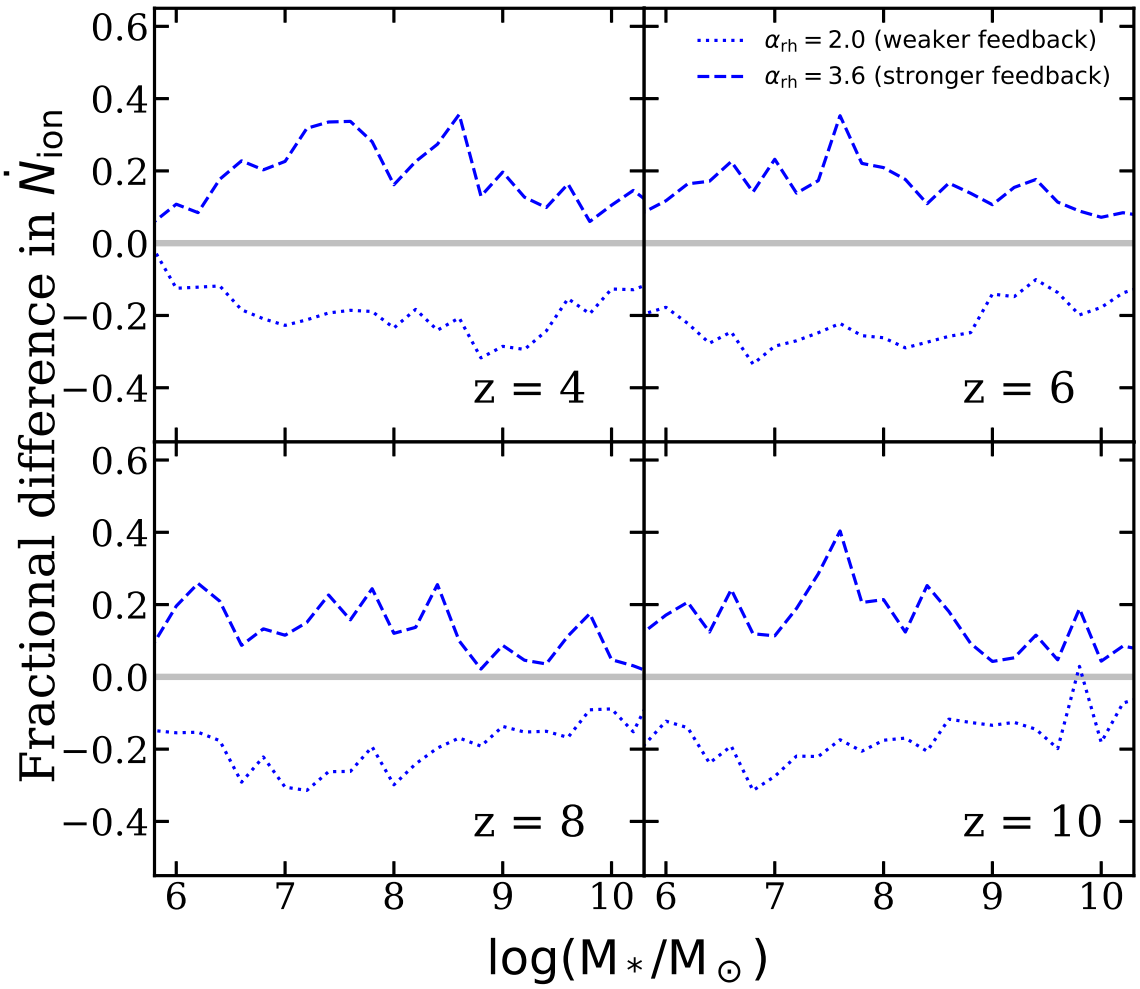

Figure 4.11 The fractional difference of the median of the distribution of $\xi_{\rm ion}$ versus stellar mass for different SN feedback slopes predicted at $z = 4$, 6, 8, and 10 with other model parameters and ingredients left at their fiducial values. The dashed and dotted blue lines represent the stronger and weaker feedback scenarios.

and observed frame magnitude. We also take advantage of the high efficiency of our model to explore how these predictions may be impacted by modelling uncertainties regarding SN feedback.

4.4.1 Comparison with observations

Our predictions for $\xi_{\rm ion}$ are in good agreement with most of the observational measurements at $4 < z < 5$, but despite the range of models we investigated in this work, none of our models reproduce the observed values of $\xi_{\rm ion}$ at $z > 5$. Our models tend to predict values of $\xi_{\rm ion}$ at $z \sim 7$ that are lower than the observational estimates by about 0.3 dex. Here we briefly discuss a few reasons that could give rise to this discrepancy. First, inferring $\xi_{\rm ion}$ from observations is highly non-trivial. This estimate relies on obtaining the intrinsic rest-frame $M_{\rm UV}$ using the UV slope $\beta_{\rm UV}$ to correct for dust extinction. However, it is not known how reliable the empirical relations between UV slope and attenuation are at these redshifts, and uncertainties in the underlying shape of the attenuation curve also lead to uncertainties (Bouwens et al. 2016c).

In addition, due to limitations in current survey instruments, objects with extremely bright emission lines are preferentially selected for follow-up. These may be biased towards objects at the extreme tails of the distribution of physical properties, or towards rare starbursting galaxies. The large sample of lower redshift galaxies ($0 < z < 1$) studied by Izotov et al. (2017) show a broad range of values of $\xi_{\rm ion}$, and suggest the strong sensitivity in this parameter to recent bursts.

Finally, the estimates of $\xi_{\rm ion}$ from the stellar population synthesis models are still quite uncertain. For example, the predicted values of $\xi_{\rm ion}$ dropped by about 0.1 to 0.2 dex following changes to the stellar atmosphere models implemented in v2.1 of the BPASS models. However, Eldridge et al. (2017) note that an improved treatment of rotation could increase $\xi_{\rm ion}$ by a similar amount (see also Levesque et al. (2012)). Furthermore, $\xi_{\rm ion}$ is quite sensitive to the high-mass end of the stellar IMF. For example, Eldridge et al. (2017) show that increasing the upper mass cut-off of the IMF from 100 to 300 $M_\odot$ increases the value of $\xi_{\rm ion}$ by about 0.15 dex in the v2.1 BPASS binary models.

4.4.2 Our results in the context of other theoretical studies

SAMs and hydrodynamic simulations are two different approaches to modelling galaxy formation (see Somerville & Davé (2015) for a detailed discussion), but they both require the use of phenomenological recipes for processes that cannot be explicitly simulated, such as star formation and stellar feedback. Both our models and the BlueTides simulations adopted similar multi-phase gas-partitioning and H_2-based star formation recipes, however, a number of other differences in the models may contribute to the discrepancies in their predictions. These include the choice of IMF, the quantities given priority in the calibration process, and the set of cosmological parameters adopted.

BlueTides adopts cosmological parameters that are consistent with measured values reported by WMAP, which are different from the ones adopted in this work. However, we do not expect this to lead to major differences in the relevant predicted galaxy properties (see discussion in Appendix C in Chapter 3). In the model comparison presented in Chapter 2 and Chapter 3, we showed that the predicted statistical properties (rest-UV LFs, stellar mass functions, SFR functions) of galaxies at $z \sim 8 - 10$ are nearly identical in our models and in BlueTides. However, when we compared our predicted values of $\xi_{\rm ion}$ at a given stellar mass at $z \sim 8$ with those predicted by W16 for BlueTides, we find that the values agree well for the BC03 SPS models, while our predicted values are lower by about 0.1 dex for the single star bpass models and about 0.2 dex lower for the bpass models that include binary star evolution. We argue, based on the results presented in Eldridge et al. (2017), that this difference is largely accounted for by changes in the bpass model results from the v2.0 version used by W16 to the v2.2.1 version that we have adopted in this work. This is supported by the fact that the predicted value of $\xi_{\rm ion}$ from the FirstLight simulations (which also adopted the bpass v2.2.1 models) are quite similar to ours.

When we compare the other stellar population properties of galaxies in BlueTides and our models at the same stellar mass and redshift ($z = 8$), we find that our galaxies are slightly younger and have higher specific star formation rates. By far the most significant difference, however, is that the stellar populations have mass-weighted metallicities that are a factor of six higher in our models than in BlueTides.

All of the physically grounded models in the literature (including BlueTides, FirstLight, and our models) predict rather mild redshift evolution in $\xi_{\rm ion}$ as well as generally lower values than those implied by observations at $z \gtrsim 5$.

4.4.3 Caveats and limitations of the modelling framework

The productivity of ionizing photons of high-redshift galaxies connects the small-scale processes of stellar evolution, star formation, and stellar feedback with cosmological scale processes such as the reionization of intergalactic hydrogen. Our results depend critically on stellar population synthesis models, which have their own limitations and uncertainties (Conroy 2013). The massive, low-metallicity stars that are the most efficient at producing ionizing photons are particularly sensitive to modelling uncertainties in physical processes, such as mass loss, mixing, convection, rotation, and magnetic fields; and significant progress has been made in the past few years on improving these models. One area that has received some attention within the literature on high-redshift galaxy populations and the epoch of reionization recently is the impact of binary processes on the evolution of massive stars and the implications for the production of ionizing photons (Wilkins

et al. 2016b; Ma et al. 2016a; Ceverino et al. 2019). Binarity can have a significant effect on the production rate of ionizing photons via a diverse range of physical mechanisms. Mass transfer and mergers among binaries can increase the number of massive stars present over times longer than the lifetime of massive single stars. In addition, mass transfer can produce rapidly rotating 'chemically homogeneous' stars, which are more prevalent at low metallicity and are highly efficient at producing ionizing photons. The rotating massive stellar population can be an significant source of ionizing photons in the first ~ 4 Myr for starburst galaxies (Choi et al. 2017b). A third process is that stars in a binary system can be stripped of their envelope, exposing the hot core of the star. Stripped stars can boost the production rate of ionizing photons at times 20–500 Myr after the onset of star formation by as much as 1–2 orders of magnitude Götberg et al. (2019). The BPASS models adopted in this work include modelling of these processes (Eldridge et al. 2011; Eldridge & Stanway 2012; Stanway et al. 2016; Eldridge et al. 2017), while the role of stripped stars has been highlighted in Götberg et al. (2019). Other processes that may be important for producing ionizing photons, but which are not accounted for or not modelled in as much detail in the BPASS models, include accreting white dwarfs and X-ray binaries (Madau & Fragos 2017). Because $\xi_{\rm ion}$ is so sensitive to the most massive stars, it is also quite sensitive to details of the high-mass slope and cut-off of the stellar IMF (e.g. Eldridge et al. 2017). In summary, the raw predictions of $\xi_{\rm ion}$ even in the latest models seem to be uncertain at the level of at least a factor of two.

Moving up in scale, the star formation and chemical enrichment histories that form the backbone of this work are based on the modelling framework developed in a series of papers including SPT15, Chapter 2, and Chapter 3. These models make a large number of simplifying assumptions about how baryonic processes in galaxies operate. A fundamental ansatz of our models is that the scaling relations that describe processes such as gas partitioning, star formation efficiency, and stellar feedback are valid over all of cosmic time. This may not necessarily be the case. Our adopted scaling relations may have dependencies that do not properly track the true physical properties, which may evolve differently. These assumptions are typically tested and calibrated by comparing with observations. As shown in these previous works, the predicted distribution functions for $M_{\rm UV}$, M_*, and SFR up to $z \sim 10$ are in very good agreement with observations in regimes where observational constraints are available. However, the physical properties of high-redshift faint galaxies, as well as the underlying physical processes that drive them, are poorly constrained.

The results of this work hint that the treatment of stellar feedback and chemical evolution may be an area that requires careful examination in future work. Although our models produce excellent agreement with both the BlueTides and FIRE simulations with regard to bulk properties such as the stellar-to-halo-mass ratio, as noted above, the predicted mass-metallicity relation at $z \sim 6$–8 is

substantially different. As shown in figure 7 of Ma et al. (2016b) and figure 6 of Somerville & Davé (2015), the predicted mass-metallicity relation shows a large dispersion between different models and simulations. This is because there are many details of the chemical enrichment process that are poorly understood or simplified in cosmological galaxy formation models. For example, the version of the SC-SAM models used here assumes instantaneous recycling, i.e., that metals ejected by SNe are instantaneously mixed with the whole ISM. Our models assume a fixed chemical yield, y, which is treated as a fixed parameter that is tuned such that the predicted stellar mass-metallicity relation matches $z \sim 0$ observations. We also assume that the metallicity of outflows is the same as that of the ISM, while there is evidence in nearby galaxies that stellar driven winds may be substantially metal-enhanced (metals preferentially ejected Chisholm et al. 2018). Furthermore, we note that the multiphase-gas partitioning recipes adopted in our models may break down in extremely metal-poor environments (Sternberg et al. 2014). Neither primordial H_2 cooling and Pop III stars nor metal enrichment by these objects is explicitly included in our models. Instead, top-level progenitor halos are polluted to a metallicity floor with typical values of $Z_{\text{pre-enrich}} = 10^{-3} Z_{\odot}$.

4.4.4 Characterizing ionizing sources with JWST and beyond

As shown in Fig. 4.9, predictions from our model show that β_{UV} may not be an very good tracer for ξ_{ion} as the correlation between quantities is rather weak. The productivity of ionizing photons can be indirectly traced with a number of spectral features. Due to current instrumental limitations, it is tremendously challenging to obtain high-resolution spectra for galaxies at high redshifts. The NIRSpec onboard *JWST* possesses the advanced capability for detecting Hα at $z \lesssim 7$, and given adequate exposure time, the Mid-Infrared Instrument (MIRI) will also be able to pick up signals from bright objects at $z \gtrsim 7$. Furthermore, planned flagship ground-based telescopes, such as the Extremely Large Telescope (ELT), the Thirty Meter Telescope (TMT), and the Giant Magellan Telescope (GMT), are also expected to enable these kinds of observations for many more objects and for fainter, perhaps more 'typical' galaxies. These anticipated observations will be able to constrain or even rule out completely some of the predictions provided in this work.

4.4.5 Implications of our results for cosmic reionization

One of the main goals of this work is to provide a physical explanation for how the intrinsic production efficiency of ionizing photons scales with physical properties of galaxies and how these relations have evolved during the EoR. A very common assumption in reionization modelling (e.g. Robertson et al.

2013) is that the cosmic ionizing emissivity can be written as

$$\dot{n}_{\rm ion} = f_{\rm esc}\xi_{\rm ion}\rho_{\rm UV} \tag{4.3}$$

i.e., implicitly assuming that $f_{\rm esc}$ and $\xi_{\rm ion}$ have a fixed value with no scatter, that does not change over cosmic time. While this may be true in an effective, integrated sense, in this thesis, we have developed more realistic models in which both of these parameters may be correlated with internal galaxy properties, such as their UV luminosity, stellar mass, or metallicity. These properties, in turn, are expected to be correlated with galaxy clustering and large scale environment, which likely has implications for reionization. We are building towards a model framework that can efficiently go from dark matter properties and a set of assumptions about the physical processes involved in galaxy formation, to a cosmic reionization history and ultimately to predicted maps of observable tracers of the reionization.

A summary of our predictions for the *intrinsic* cosmic ionizing emissivity is shown in Fig. 4.12. This figure illustrates the impact of adopting different stellar population models (self-consistently incorporated within our models) and of varying the strength of stellar feedback. As discussed above, the rather strong metallicity dependence of $\xi_{\rm ion}$ implies that the details of the interplay between chemical evolution and stellar populations can lead to differences of up to factors of a few, which could be significant for reionization. The total number of ionizing photons produced are degenerately affected by the abundance of galaxies during the EoR and the intrinsic productivity of ionizing radiation. As shown in Chapter 2 and Chapter 3, the abundances of low-mass galaxies are very sensitive to the strength of SN feedback and the number of objects predicted can span about an order of magnitude for different model assumptions. On the other hand, in §3.4 we also show that stronger feedback yields slightly ($\sim 20\%$) boosted ionizing photon production efficiencies.

Whether galaxies alone are sufficient to ionize the IGM remains a fundamental open question. Historically, there has been tension between the total ionizing photon budget accounted for by galaxies, and constraints on the onset and duration of the EoR. Analytic calculations (e.g. Kuhlen & Faucher-Giguère 2012; Robertson et al. 2015) have shown that galaxies are likely to have produced sufficient numbers of ionizing photons to reionize the Universe under the condition that faint objects remain fairly numerous below the current detection limit and are fairly efficient at producing ionizing photons. In Chapter 2 and Chapter 3, we showed that in our models, the faint-end slope of the rest-frame UV LF remains steep down to $M_{\rm UV} \sim -9$, under the assumption that SNe feedback at high redshifts operates with similar efficiency as that required to match $z \sim 0$ observations. The predicted ionizing photon production efficiencies, especially with models that include binary

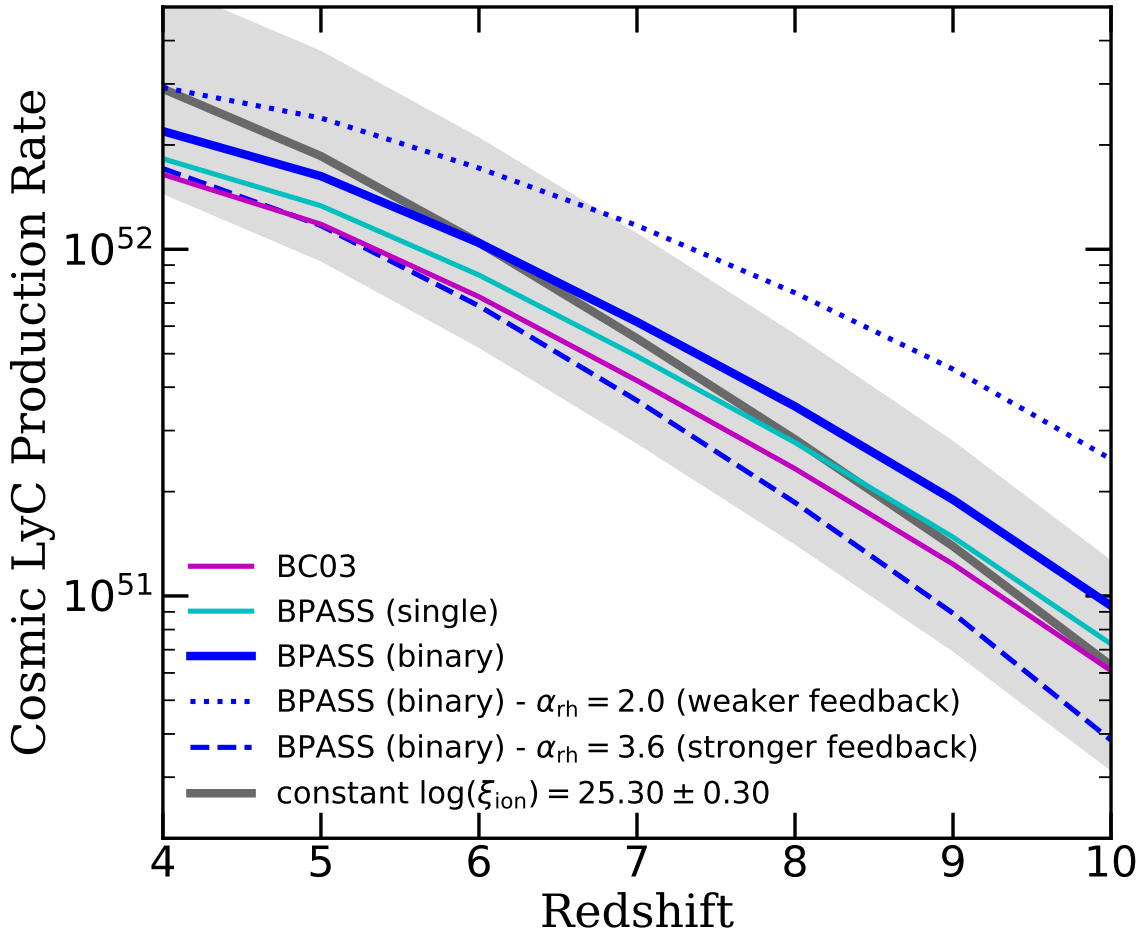

Figure 4.12 Cosmic LyC photon production rate evolution with redshift given in units of Mpc^{-3} s^{-1}, calculated using SPS models from BC03 (purple) and BPASS (cyan and blue for single and binary). For different SN feedback slopes predicted at $z = 4 - 10$ with other model parameters and ingredients left at their fiducial values, the dashed and dotted blue lines represent the stronger and weaker feedback scenarios, respectively. The grey lines show the case where $\log(\xi_{\rm ion}) = 25.30 \pm 0.30$.

evolution, are well within the range $\log(\xi_{\rm ion}) \sim 25.30 \pm 0.30$ (or higher) adopted in recent models that are able to plausibly account for observable constraints on the reionization history (Robertson et al. 2013, 2015; Kuhlen & Faucher-Giguère 2012; Finkelstein et al. 2019). Therefore, we can already anticipate that our models will also produce good agreement with the observed reionization history, subject to uncertainties on the escape fraction of ionizing radiation. We present a detailed analysis of the implications for the cosmic reionization history of our full modelling framework, including the escape fraction, in Chapter 5.

4.5 Summary and Conclusions

In this work, we implemented a physically motivated approach to calculating $\dot{N}_{\rm ion}$ in the Santa Cruz SAM, accounting for stellar age and metallicity distribution of the stellar population in each galaxy. We present results predicted based on SSP models from BC03 and both single and binary stellar population by BPASS. Taking advantage of the high efficiency of our modelling framework, we make predictions for galaxies across $z = 4$–10, forming in a wide range of halo masses, ranging from $V_{\rm vir} \approx 20$–$500\,{\rm km\,s^{-1}}$. These predictions provide forecasts for future observations with *JWST* and also put modelling constraints on the role high-redshift, low-mass galaxies played throughout the EoR. We also provide predicted scaling relations of $\dot{N}_{\rm ion}$ with M_*, $M_{\rm h}$, $\beta_{\rm UV}$, and other galaxy properties. We compared our predictions to other models from the literature and the latest observations in both the intermediate- and high-redshift universe.

We summarize our main conclusions below.

1. In agreement with previous work, we find that SPS models accounting for binary stellar populations can produce a factor of ~ 2 more ionizing radiations than models that only account for single star populations for galaxies across all mass ranges.

2. We find that faint, low-mass galaxies can have ionizing efficiencies that are up to a factor of two higher than those of bright, massive ones. This is due to the strong correlation between stellar mass (or luminosity) and metallicity that is already in place at early times in our models, and the strong dependence of $\xi_{\rm ion}$ on metallicity, particularly in models that include physical processes related to stellar binaries.

3. Our models predict a rather weak dependence of $\xi_{\rm ion}$ with redshift (about ~ 0.23 dex or factor of ~ 0.6 decrease from $z \sim 12$ to 4).

4. We find that increasing of the strength of SN feedback may allow galaxies to produce $\sim 20\%$ more ionizing photons, because their star formation is more bursty. However, we also note that stronger feedback suppresses star formation in low-mass halos and decreases the number density of low-mass galaxies, which ultimately decreases the total number of ionizing photons.

5. Our predicted median values of $\xi_{\rm ion}$ are significantly lower than all current observational estimates at $z > 5$. This may be due to sampling bias towards extremely bright and bursty objects in current high-redshift observations, or to modelling uncertainties in stellar population synthesis models.

6. We provide scaling relations for $\dot{N}_{\rm ion}$ as a function of $M_{\rm h}$ and a number of other galaxy

properties, which may be useful for cosmological scale reionization simulations where baryonic physics is not resolved.

Chapter 5

Implications for Cosmic Reionization and LyC Escape Fraction

Galaxies forming in low-mass halos are thought to be primarily responsible for reionizing the Universe during the first billion years after the Big Bang. Yet, these halos are extremely inefficient at forming stars in the nearby Universe. In this work, we address this apparent tension, and ask whether a physically motivated model of galaxy formation that reproduces the observed abundance of faint galaxies in the nearby Universe is also consistent with available observational constraints on the reionization history. By interfacing the Santa Cruz semi-analytic model for galaxy formation with an analytic reionization model, we constructed a computationally efficient pipeline that connects 'ground-level' galaxy formation physics to 'top-level' cosmological-scale observables. Based on photometric properties of the galaxy populations predicted up to $z = 15$, we compute the reionization history of intergalactic hydrogen. We quantify the three degenerate quantities that influence the total ionizing photon budget, including the abundance of galaxies, the intrinsic production rate of ionizing photons, and the LyC escape fraction. We explore covariances between these quantities using a Markov chain Monte Carlo method. We find that our locally calibrated model is consistent with all currently available constraints on the reionization history, under reasonable assumptions about the LyC escape fraction. We quantify the fraction of ionizing photons produced by galaxies of different luminosities and find that the galaxies expected to be detected in *James Webb Space Telescope* Near-Infrared Camera (NIRCam) wide and deep surveys are responsible for producing $\sim$ 40–80% of ionizing photons throughout the Epoch of Reionization.

This chapter is reproduced from published work *Semi-analytic forecasts for JWST – IV. Implications for cosmic reionization and LyC escape fraction* (Yung et al. 2020. MNRAS, 496, 4574).

5.1 Introduction

During the Epoch of Reionization (EoR), the intergalactic medium (IGM) underwent a global phase transition, during which the hydrogen progressively became ionized by the radiating Lyman-continuum (LyC) sources in the early universe (Miralda-Escude, Haehnelt, & Rees 2000). Identifying

and characterizing these sources remains a fundamental open challenge in modern cosmology. Indeed, this is one of the main science drivers of the *James Webb Space Telescope* (*JWST*). With the unprecedented infrared (IR) sensitivity and resolution of its on-board photometric instrument Near-Infrared Camera (NIRCam), *JWST* is expected to detect many more faint galaxies during the EoR. In addition, *JWST* will be able to provide additional constraints on the nature of the sources that reionized the Universe, such as revealing early accreting black holes. A number of planned *JWST* observations, including both Guaranteed Time Observation (GTO), such as the *JWST* Advanced Deep Extra-galactic Survey (JADES; Williams et al. 2018) and Early Release Science (ERS) projects, such as the Cosmic Evolution Early Release Science survey (CEERS; Finkelstein et al. 2017), are designed to study and put constraints on the galaxy populations during the EoR, including both their statistical properties and the production rate of ionizing photons.

5.1.1 The overall budget of ionizing photons

It is clear that galaxies forming in the early universe have influenced large-scale events (Dayal & Ferrara 2018). The cosmic ionizing photon budget is subject to three major moving parts, including the number density of galaxies, the intrinsic productivity of ionizing photons, and the LyC escape fraction. The volume-averaged number density of high-redshift galaxies is only partially constrained by existing observations, which are limited by the sensitivity of existing facilities, particularly in the observed frame near-mid IR. To date, nearly 2000 galaxy candidates at $z \gtrsim 6$ have been detected in the Cosmic Assembly Near-IR Deep Extragalactic Legacy Survey (CANDELS; Grogin et al. 2011; Koekemoer et al. 2011), *Hubble* Ultra Deep Field (HUDF; Beckwith et al. 2006; Bouwens et al. 2011; Ellis et al. 2013; Oesch et al. 2013), and UltraVISTA (McCracken et al. 2012), with faint objects of rest-frame UV luminosities reaching $M_{\rm UV} \sim -17$ (e.g. Bouwens et al. 2015b; Finkelstein et al. 2015a). Lensed surveys through massive foreground clusters can reach even fainter detection limits (e.g. Livermore et al. 2017; Lotz et al. 2017; Atek et al. 2018), though this approach comes with high systematic uncertainties that remain poorly constrained (Kawamata et al. 2016; Bouwens et al. 2017; Priewe et al. 2017). As a result, there are still significant uncertainties on the faint-end slope of the UV luminosity functions (UV LFs) at $z \gtrsim 6$, which give rise to uncertainties of $\gtrsim 0.2$ dex on the integrated UV luminosity density at high redshift (Ishigaki et al. 2018), as well as the magnitude at which the UV LFs 'turnover'.

The intrinsic production efficiency of ionizing radiation of high-redshift galaxies is subject to its own set of uncertainties. In early analytic calculations, this quantity was treated simply as a constant or as a parametrized function of redshift (Madau et al. 1999; Finkelstein et al. 2012a; Kuhlen & Faucher-Giguère 2012; Robertson et al. 2015; Mutch et al. 2016). However, it is now recognized

that this quantity depends strongly on many properties of the stellar populations in these early galaxies, including age, metallicity, upper mass cutoff of the stellar initial mass function (IMF), and binarity (Eldridge & Stanway 2009; Eldridge et al. 2017; Topping & Shull 2015; Wilkins et al. 2016b; Yung et al. 2020b). There are still significant uncertainties in predictions of this quantity even in state-of-the-art stellar population synthesis (SPS) models (Conroy 2013). In general, we expect high-redshift galaxies to have younger, lower metallicity stellar populations, resulting in harder spectra yielding higher LyC production efficiencies. The contribution to the ionizing photon budget from sources such as X-ray binaries and Active Galactic Nuclei (AGN) also remains uncertain (e.g. Madau & Fragos 2017; Manti et al. 2017). Some recent studies have set out to constrain the production efficiency both locally and at high redshift using observations of UV-continuum slope, $\beta_{\rm UV}$, Hα and CIV emission (Stark et al. 2015; Bouwens et al. 2016c; Schaerer et al. 2016; Shivaei et al. 2018; Emami et al. 2019; Lam et al. 2019).

The fraction of ionizing radiation escaping to the IGM is the least constrained component among these three moving parts. Simulations have shown that it is extremely sensitive to many detailed geometrical and physical features that act across many scales, including the internal distributions of dense gas, dust clouds, and stars within the interstellar medium (ISM) and the structure of the circumgalactic medium (CGM) (Paardekooper et al. 2011, 2013, 2015; Benson et al. 2013; Kimm & Cen 2014; Kimm et al. 2017, 2019; Ma et al. 2015; Xu et al. 2016; Popping et al. 2017a; Trebitsch et al. 2017, 2018). These studies also found that the escape fraction does not correlate well with any particular global physical galaxy property and can scatter across an extremely wide range, from less than a thousandth to a few tens of a percent, even for galaxies of similar physical properties forming at the same epoch. Many studies have attempted to constrain the escape fraction via observations and arrived at similar conclusions (e.g. Vanzella et al. 2010, 2015, 2018; Dijkstra et al. 2016; Guaita et al. 2016; Shapley et al. 2016; Grazian et al. 2017; Fletcher et al. 2019; Nakajima et al. 2020). Similar to the LyC production rate, many previous studies have treated the escape fraction as a single value (Finkelstein et al. 2010, 2012a, 2015a; Robertson et al. 2013, 2015; Bouwens et al. 2015b) or as a parametrized function of redshift or galaxy physical properties (Wyithe et al. 2010; Kuhlen & Faucher-Giguère 2012; Sharma et al. 2016; Naidu et al. 2020; Finkelstein et al. 2019).

It is clear that detected galaxies alone are far from sufficient to reionize the universe (Madau et al. 2008; Finkelstein et al. 2015a; Robertson et al. 2015). However, by assuming a LyC production efficiency and escape fraction that is consistent with that of bright galaxies, analytic calculations have shown that faint galaxy populations extrapolated from the observed UV LFs to below the current detection limits are able to provide the amount of ionizing photons needed to fully reionize

the Universe in the required time-frame (Finkelstein et al. 2012a, 2015a, 2019; Kuhlen & Faucher-Giguère 2012; Robertson et al. 2015; Bouwens et al. 2015a; Stark 2016).

5.1.2 Constraints on the Epoch of Reionization

The reionization history of the intergalactic hydrogen is constrained by a variety of IGM and CMB observations (Fan et al. 2006a). During the phase transition, the depletion of neutral hydrogen along the line of sight can partially absorb high-redshift quasar spectra and leave behind a feature known as the Gunn-Peterson Trough (Gunn & Peterson 1965; Becker et al. 2001; Fan et al. 2006b). The presence of intervening HI also decreases the visibility of Lyα emitters, which puts a lower-limit to the redshift of the onset of the EoR (Stark et al. 2010; Dijkstra et al. 2011; Pentericci et al. 2011, 2014; Schenker et al. 2012, 2014; Treu et al. 2013; Tilvi et al. 2014; Schmidt et al. 2016; Mason et al. 2018a). This same mechanism also enables the 'Lyman-break selection' technique for identifying high-redshift galaxy candidates (Steidel et al. 1996, 1999). On the other hand, the CMB is scattered and polarized by free electrons in an ionized IGM. Therefore, the measured Thomson optical depth of the CMB, $\tau_{\rm CMB}$, can be used to constrain the total number of electrons along the line of sight to the IGM. The neutral IGM fraction towards the end of the EoR is constrained by a variety of observations (see Robertson et al. 2015 for a concise summary). Combining constraints on the onset and duration of the reionization process from various observations, the astronomical community has come to a general consensus that the phase transition of intergalactic hydrogen occurred approximately between $z = 6$–10, and this period is often referred to as the Epoch of Reionization (EoR).

Historically, there has been tension among different observational constraints on the onset and duration of reionization. Early measurements of $\tau_{\rm CMB}$ reported by the Cosmic Background Explorer (COBE; Kamionkowski et al. 1994) and the Wilkinson Microwave Anisotropy Probe (WMAP; Spergel et al. 2003, 2007; Komatsu et al. 2009, 2011; Hinshaw et al. 2013) seemed to imply a rapid reionization with a rather early conclusion (e.g. Somerville & Livio 2003; Kuhlen & Faucher-Giguère 2012; Robertson et al. 2013). On the other hand, a collection of Lyα forest constraints indicates that the number of ionizing photons reaching the IGM gradually flattens or even declines at $z \sim 2$–6 (Bolton & Haehnelt 2007; Faucher-Giguère et al. 2008a, Prochaska, Worseck, & O'Meara 2009, Songaila & Cowie 2010). It was difficult to reconcile the early reionization apparently implied by the CMB (requiring a certain budget of ionizing photons) with the rather low emissivity at $z \sim 4$–6, while the galaxy population had presumably grown. One way to reconcile this tension was by invoking an 'exotic' population of ionizing sources that contributed only at high redshift (such as Pop III stars or mini-quasars) or an escape fraction that strongly decreased with cosmic time. However,

recent estimates of $\tau_{\rm CMB}$ reported by the Planck Collaboration (2014; 2016a; 2018) have become considerably lower, indicating later reionization. At the same time, more recent work on the cosmic emissivity from Lyα forest constraints by Becker & Bolton (2013) indicates a higher emissivity towards the end of EoR at $z \sim 4$–6, largely alleviating the tension. However, these measurements still provide important complementary constraints on the reionization history.

Another puzzle that has been discussed is the potential tension between the apparent need for relatively *efficient* star formation in low mass halos at high redshift, needed to supply adequate numbers of the faint, low-mass galaxies that are invoked to make up the shortfall in the ionizing photon budget, and the much more *inefficient* star formation in low-mass halos required to reconcile observed galaxy luminosity functions at low redshift with predicted halo mass functions in Λ Cold Dark Matter (Lu et al. 2014b; Madau & Dickinson 2014). Observations of faint, low-mass galaxies in the nearby Universe provide important complementary constraints to deep-field studies on EoR populations (Weisz et al. 2014; Boylan-Kolchin et al. 2014, 2015, 2016; Graus et al. 2016).

5.1.3 Current simulation efforts

Modelling cosmic reionization is extremely challenging because, as we have outlined, it depends on accurately simulating structures from sub-pc scales to the largest structures in the Universe (~ 100 Mpc). Several different complementary approaches have been presented in the literature. High-resolution cosmological zoom-in simulations, such as *Renaissance* (Wise et al. 2012; O'Shea et al. 2015), FIRE (Hopkins et al. 2014, 2018; Ma et al. 2018a,b) and FirstLight (Ceverino et al. 2019) simulate small volumes at relatively high resolution. They are able to study the detailed properties of galaxies and their ISM, down to scales of tens of pc, but it is not feasible to simulate large volumes. Larger volume numerical hydrodynamic simulations such as EAGLE (Schaye et al. 2015), Illustris and IllustrisTNG (Genel et al. 2014; Pillepich et al. 2018), CROC (Gnedin 2014; Gnedin & Kaurov 2014; Gnedin 2016; Gnedin et al. 2017), CoDa (Ocvirk et al. 2016, 2020), BlueTides (Feng et al. 2016; Wilkins et al. 2017), Simba (Davé et al. 2019; Wu et al. 2020), and Sphinx (Rosdahl et al. 2018) are able to simulate larger volumes ($\sim$ 10–100 Mpc), but may not resolve the very low-mass halos that could be important for reionization, or the detailed properties of the ISM. A third approach is to simulate large volumes, in some cases with explicit modelling of radiative transfer in the IGM, but treating sources in a simplified way, e.g. by adopting empirical relations relating rest-UV luminosity to halo virial mass or stellar mass to estimate the number of ionizing photons produced (Iliev et al. 2006a,b; Trac & Cen 2007; Trac & Gnedin 2011; Santos et al. 2010; Hassan et al. 2016, 2017). This approach essentially operates under the same guiding principle that drives the popular (semi-)empirical modelling approach (Behroozi et al. 2019; Moster et al. 2018; Tacchella

et al. 2018; Finkelstein et al. 2019). However, this relies strongly on observational constraints, which must be extrapolated in regimes where these relations are not well calibrated.

The semi-analytic modelling approach is a middle way of bridging the gap between galaxy formation physics and the large-scale reionization history using physically motivated relationships between dark matter halo formation histories and galaxy properties. Semi-analytic models have had a long history of contributing to advancing the understanding of galaxy formation in ways that are complementary to numerical simulations (Cole et al. 1994; Kauffmann et al. 1993; Somerville & Primack 1999; Somerville & Davé 2015). They are grounded in the framework of dark matter halo 'merger trees', and adopt simplified but physically motivated analytic recipes to model the main processes that shape galaxy formation. The models contain phenomenological parameters that are calibrated to reproduce a set of key observational relations in the nearby Universe. The models that we adopt here, the Santa Cruz semi-analytic models (Somerville & Primack 1999; Somerville et al. 2008, 2015), have also been shown to reproduce a broad suite of other observations over a range of cosmic time and galaxy mass. The semi-analytic approach to studying reionization has also been adopted by the DRAGONS project (Liu et al. 2016; Mutch et al. 2016; Geil et al. 2016). Because of the computational efficiency of the semi-analytic approach, we are able to simulate large volumes down to the lowest mass halos that are expected to be able to cool via atomic cooling. In addition, we are able to explore variations in our model parameters. We have compared our model predictions with those from both high-resolution and large-volume cosmological hydrodynamic simulations (Chapter 2 and Chapter 3), and find excellent agreement.

The key components of this chapter are summarized as follows: the semi-analytic modelling pipeline, including the Santa Cruz galaxy formation model and the analytic reionization model are summarized briefly in Section 5.2. Predicted reionization histories along with exploration of the effect of varying different model components are presented in Section 5.3, including some specific predictions regarding *JWST* in Section 5.3.4. We discuss our findings in Section 5.4, and a summary and conclusions follow in Section 5.5.

5.2 The Modelling Framework

In this section, we present the components of a joint semi-analytic modelling pipeline for galaxy formation and cosmic reionization used to carry out this study. Throughout this work, we adopt cosmological parameters that are consistent with the ones reported by Planck Collaboration in 2015: $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $H_0 = 67.8\,{\rm km\,s^{-1}Mpc^{-1}}$, $\sigma_8 = 0.831$, and $n_s = 0.9665$. We adopt hydrogen and helium mass fractions $X = 0.75$ and $Y = 0.25$.

5.2.1 Semi-analytic model for galaxy formation

The galaxy populations that source the ionizing photons are predicted using a slightly modified version of the well-established Santa Cruz SAM outlined in Somerville, Popping, & Trager (2015, hereafter SPT15). We refer the reader to the following works for full details of the modelling framework: Somerville & Primack (1999); Somerville, Primack, & Faber (2001); Somerville et al. (2008); Somerville et al. (2012); Popping, Somerville, & Trager (2014, hereafter PST14) and SPT15. For details on the model parameters used in this thesis and how they were calibrated, see Appendix B.

The semi-analytic approach of modelling galaxy formation is based upon the merger histories of dark matter halos, sometimes referred to as 'merger trees'. In this work, we adopted merger trees that are constructed using the Extended Press-Schechter (EPS) formalism (Press & Schechter 1974; Lacey & Cole 1993; Somerville & Kolatt 1999), which have been shown to well-reproduce the statistical results for a large ensemble of merger trees extracted from N-body simulations (Somerville & Kolatt 1999; Somerville et al. 2008; Zhang et al. 2008; Jiang & van den Bosch 2014). This approach is able to achieve a wider dynamic range than any existing cosmological simulations, while requiring only a small fraction of computation resources. For these reasons, our physical models are able to account for halos ranging from the very low-mass ones near the atomic cooling limit to the rare, massive ones across a wide range of redshift. The number density of 'root' halos is computed based on results cosmological dark matter simulations (Klypin et al. 2016; Rodríguez-Puebla et al. 2016; Visbal et al. 2018). For further details, see Chapter 3.

Within these merger trees, SAMs then implement a set of coupled ordinary differential equations describing the flow of mass and metals between different components (diffuse intergalactic gas, hot halo gas, cold interstellar gas, the stellar body of the galaxy, etc). These flows are influenced by a range of physical processes, including cosmological accretion and cooling, star formation, chemical evolution, stellar-driven winds, and black hole feedback. The equations governing these processes contain "tunable" parameters that reflect our lack of a complete understanding of the basic physics. These parameters are calibrated to match a set of observational relationships at $z = 0$. Note that in this thesis, as in all previous work with the Santa Cruz SAMs, *the models have not been tuned to match observations at high redshift.*

The Santa Cruz model (PST14; SPT15) includes a multiphase gas-partitioning recipe, which subdivides the cold gas content into an atomic, ionized, and molecular component, and a H_2-based stars formation recipe, which utilizes the predicted surface density of H_2 (Σ_{H_2}) as a tracer for the

surface density of SFR ($\Sigma_{\rm SFR}$). In this work, we adopted the metallicity-based, UV-background-dependent partitioning recipe based on work by Gnedin & Kravtsov (2011, hereafter GK) and the SF relation based on observations by Bigiel et al. (2008, hereafter Big). We note that recent evidence from both theory and observation suggests that the SF relation slope may steepen to ~ 2 at higher H_2 surface densities (Sharon et al. 2013; Rawle et al. 2014; Hodge et al. 2015; Tacconi et al. 2018). In previous chapters in this thesis, we have shown that this 'two-slope' SF relation is crucial for our model to produce predicted galaxy populations that simultaneously match observational constraints on stellar mass, star formation rate, and rest-frame UV luminosity at $z = 4$–10 (Chapter 2 and Chapter 3). Thus, we refer to it as our fiducial model (GK-Big2).

The Santa Cruz SAM has been tested extensively in the past and shown to be able to reproduce a wide range of observables. In Chapter 2, the free parameters were re-calibrated to match a subset of $z \sim 0$ observations after adopting the updated cosmological parameters reported by the Planck Collaboration. We then, in Chapter 2 and Chapter 3, identified the set of physical prescriptions (e.g. SF recipes) and physical parameters (e.g. SNe feedback slope) that are required to reproduce the evolution seen in observed galaxy populations up to $z \sim 10$. This is encouraging as it suggests the physical processes that shape the formation of galaxies during reionization may not be so different from those that determine the properties of low-redshift galaxies. Taking advantage of the model's efficiency, we also quantified the impacts on the predicted galaxy populations from the uncertainties in these model components by conducting controlled experiments where we systematically varied the model parameters. We found that the key process that has strong effects on the rest-frame UV luminosities and physical properties for bright, massive galaxies is the SF efficiency or time-scale ($\tau_{*,0}$, see equation (1) in Chapter 2), which effectively characterizes the gas depletion time. For faint, low-mass galaxies, the UV LF is most sensitive to the stellar feedback relation slope ($\alpha_{\rm rh}$, see equation (3) in Chapter 2), which characterizes the dependence of the mass-loading factor of cold gas ejected by stellar feedback on halo circular velocity. Currently there are not strong constraints on the faint-end slope of the UV LFs during EoR, where the predicted number density of faint galaxies across different models can vary by up to ~ 1.5 dex.

In the following subsections, we highlight how the main moving parts affecting the total emissivity of ionizing photons are treated in this work.

Galaxy populations at ultrahigh redshifts

In order to quantify the contribution of ionizing photons from galaxies at ultrahigh redshifts ($z \gtrsim 10$), we extend the predictions from our SAM up to $z \sim 15$. To assign a volume-averaged density to these galaxies, we use the same functional form for the HMF with the fitting parameters tuned to

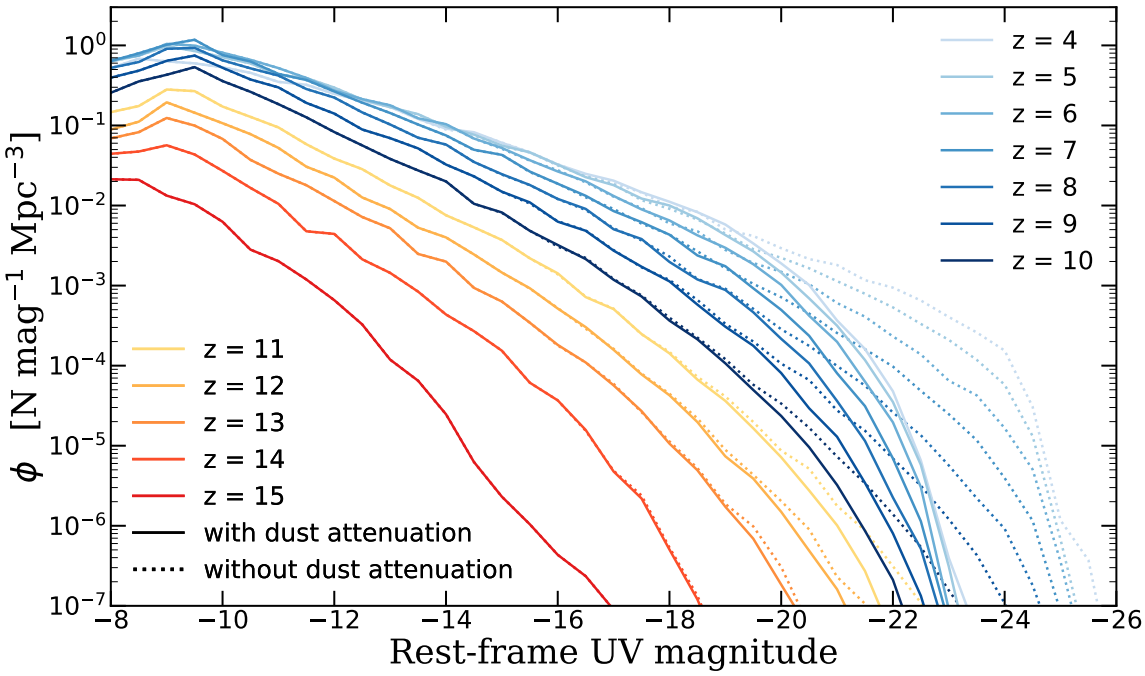

Figure 5.1 Predicted redshift evolution of the intrinsic (dotted) and dust-attenuated (solid) rest-UV LFs between $z = 11$ and 15 (this work; red colour series) and between $z = 4$ and 10 (Chapter 2; blue colour series). The turnover at the faint end is not due to resolution, but rather to the atomic cooling limit. At very high redshifts, the UVLF is not well fit by a Schechter function. At $z \sim 4-8$, the apparent knee in the observed UVLF is largely due to differential dust extinction, which is larger in more luminous galaxies.

fit the results from the same set of simulations (Bolshoi-Planck and Visbal et al., see Fig. G.1). See Appendix F for full details and the values of all parameters. In Fig. 5.1, we present both the intrinsic (dust-free) and the dust-attenuated rest-frame UV luminosity functions predicted for the extended redshift range $z = 11$–15. In the same figure, we also compare these galaxies to the evolution between $z = 4$–10 previously presented in Chapter 2. Tabulated values for the dust-attenuated UV LFs are accessible via an online repository (see Appendix A). We emphasize that the turnover at the faint end of our predicted luminosity functions is not due to resolution but is a result of the atomic cooling limit, which corresponds to a limiting halo mass that evolves with redshift. We found in our predictions that the characteristic 'knee' in UV LFs vanishes at $z \gtrsim 9$, seemingly due to both insignificant AGN feedback and lacking of dust (see fig.5.1), and the faint-end slopes also gradually flatten as a function of rest-UV. The Schechter function is no longer a good representation and therefore, we do not provide Schechter fitting parameters.

We continue to explore the impacts from modelling uncertainties in the context of cosmic reionization. In Fig. 5.2, we show UV LFs predictions for $\alpha_{\rm rh} = 2.0$, 2.4, 3.2, and 3.6. This is consistent with the findings in Chapter 2 and Chapter 3, where we showed that the faint-end slope of the UV

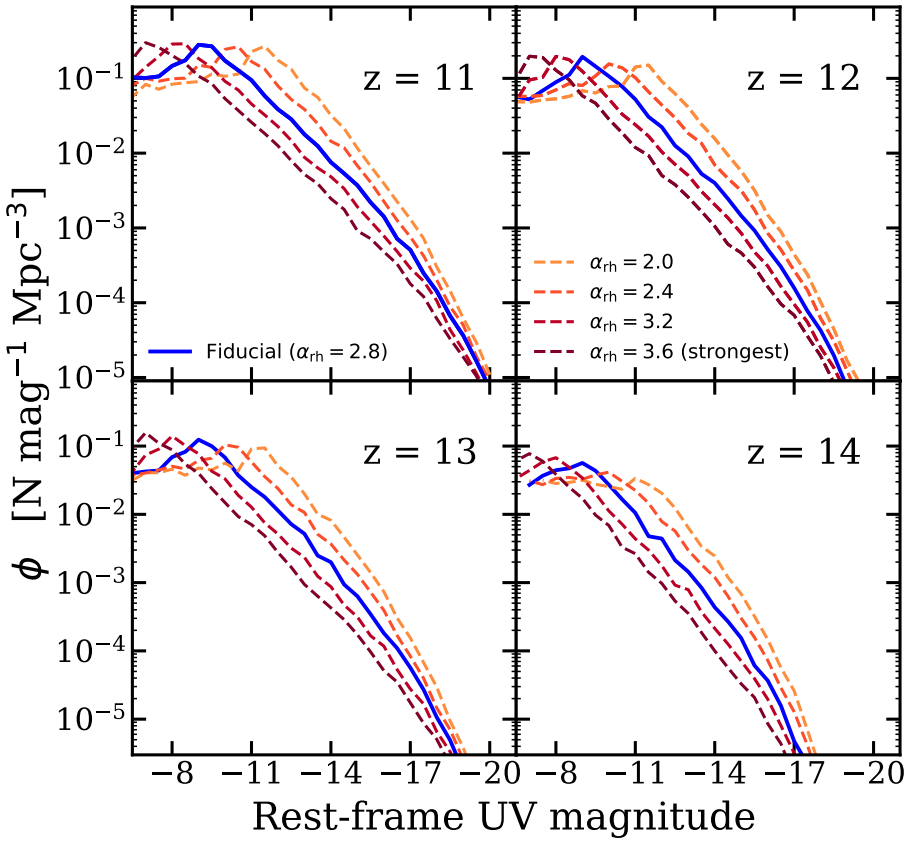

Figure 5.2 Redshift evolution of the dust-attenuated UV LFs between $z = 11$ and 14 predicted by our fiducial model ($\alpha_{\rm rh} = 2.8$, blue solid line). We show four additional scenarios where we vary the parameter controlling the mass-loading of stellar driven winds, with $\alpha_{\rm rh} = 2.0$ (weakest, lightest colour), 2.4, 3.2, and 3.6 (strongest, darkest colour). Larger values of $\alpha_{\rm rh}$ produce stronger suppression of star formation in low-mass halos, leading to shallower faint end slopes and a lower luminosity for the turnover.

LFs is inversely correlated with the stellar feedback parameter $\alpha_{\rm rh}$ (i.e., a stronger dependence of wind-mass loading on halo circular velocity leads to a flatter faint end slope). Furthermore, this effect also effectively shifts the halo occupation function and the turnover in the faint end of the UV LFs, which corresponds to the atomic cooling limit. In other words, we predict that the magnitude where the UF LF is truncated is inversely related with the strength of stellar feedback.

The intrinsic production rate of ionizing radiation

We refer to the ionizing photon production rate, $\dot{N}_{\rm ion}$, and the production efficiency, $\xi_{\rm ion}$, by stellar populations in galaxies, which does not account for the absorption or attenuation by the ISM and CGM, as 'intrinsic'. In Chapter 4, we self-consistently predict $\dot{N}_{\rm ion}$ within the Santa Cruz modelling framework, based on the predicted star formation and chemical enrichment histories and results from stellar population synthesis (SPS) models. This model component enables us to distinguish and track

the contribution from galaxies across different rest-frame UV magnitudes and stellar masses. In this work, we adopt the published results from the data tables released by the BC03[1] and the BPASS group[2] (Stanway et al. 2016; Eldridge et al. 2017; Stanway & Eldridge 2018). Both models assume a Chabrier IMF with an upper mass cutoff $m_{\rm U} = 100{\rm M}_\odot$. These predictions for $z = 4$–10 have been examined in detail in Chapter 4. In that work, we also explored the scaling relations of $\xi_{\rm ion}$ and $\dot{N}_{\rm ion}$ with many SF-related physical properties and found that $\xi_{\rm ion}$ is mildly correlated with M_* and SFR, and these scaling relations evolve mildly as a function of redshift (where the underlying driving physical parameter is predominantly stellar metallicity). Although the BPASS SPS models account for mass transfer and mergers in stellar binaries, some processes that could potentially boost the production rate of ionizing photons, such as accreting white dwarfs and X-ray binaries, are not included in these models.

Here we extend these predictions to even higher redshift galaxy populations. In Fig. 5.3, we provide the predictions for the specific ionizing photon production rate, $\dot{N}_{\rm ion}/M_{\rm h}$, for halos in a relevant mass range at $z = 11$–14. The 2D histograms are shaded according to the conditional number density (Mpc^{-1}) of galaxies in each bin, which is normalized to the sum of the number density in its corresponding (vertical) halo mass bin. The median, 16th, and 84th percentiles are marked in each panel to illustrate the statistical distribution. Comparing to the predictions between $z = 4$ to 10 shown in Fig. 4.7 in Chapter 4, which showed $\dot{N}_{\rm ion}/M_{\rm h}$ increases across the halo mass range explored, we find that the production rate per halo mass seems to have plateaued has noticeably larger scatter.

Escape fraction of Lyman-continuum photons

The LyC escape fraction can be very stochastic depending on the many intricate physical processes occurring in individual galaxies and their internal structure. In this work, we take a simplistic approach and regard it as a *population-averaged quantity*, which can either be understood as the population of galaxies all sharing the same escape fraction or as the escape fraction of the total number of ionizing photons collectively produced by all galaxies. We treat it as a controlled free parameter, which may either be a constant value or evolve as a function of redshift. For the remainder of this work, we refer to the LyC escape fraction as $f_{\rm esc}$.

Inspired by the functional form presented by Kuhlen & Faucher-Giguère (2012), we adopt the

[1] http://www.bruzual.org/ gbruzual/bc03/
[2] https://bpass.auckland.ac.nz/, v2.2.1

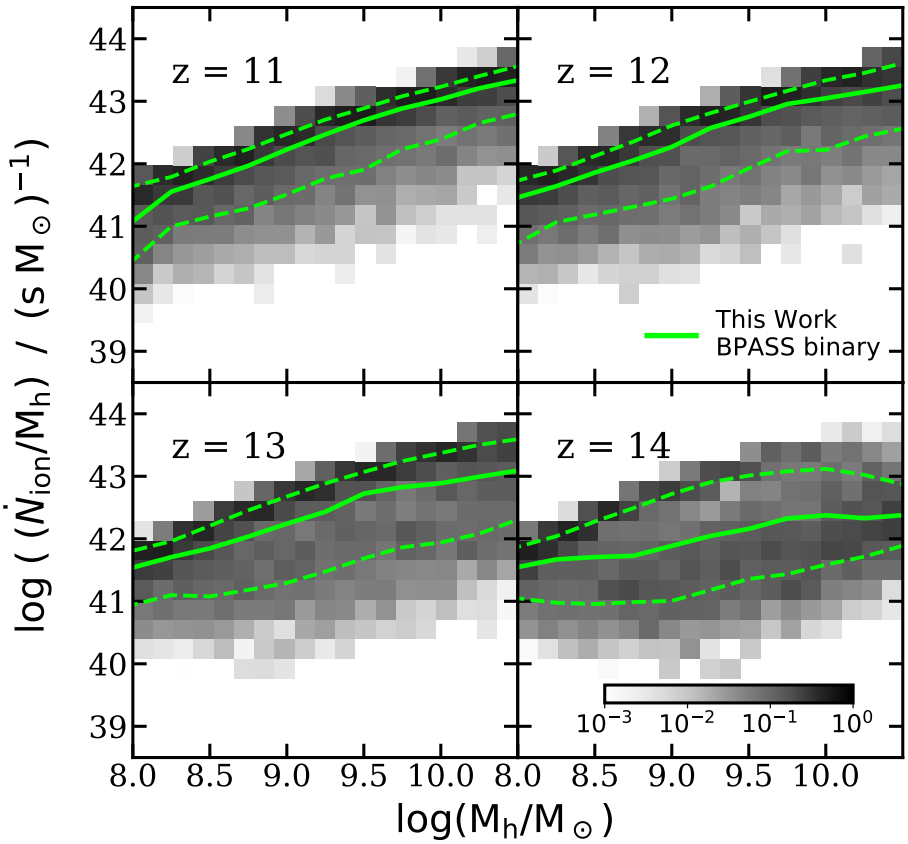

Figure 5.3 Specific ionizing photon production rate, $\dot{N}_{\rm ion}/M_{\rm h}$, as a function of halo mass between $z = 11$ and 14, predicted by our fiducial model. The green solid and dashed lines mark the 50th, 16th, and 84th percentiles. The greyscale 2D histograms show the conditional number density per Mpc^3 in each bin, normalized to the number density in the corresponding (vertical) halo mass bin. The figure shows a decline in the specific ionizing photon production rate at fixed halo mass with increasing redshift, and a flattening dependence on halo mass. This is because these very early halos have not yet had time to form many stars.

following expression for the redshift evolution of $f_{\rm esc}$:

$$f_{\rm esc}(z) = \frac{f_{\rm esc,max}}{1 + \left(\frac{f_{\rm esc,max}}{f_{\rm esc,0}} - 1\right)\, e^{-k_0(z-z_0)}}, \tag{5.1}$$

assuming $f_{\rm esc}$ decreases from some maximum value at high redshift, $f_{\rm esc,max}$, at a characteristic growth rate, k_0, until it asymptotically reaches an anchoring valuing $f_{\rm esc,0}$ at a given redshift $z_0 = 4$. A goal of this work is to obtain constraints on $f_{\rm esc}$ under this empirical parametrization, as required by the set of currently available observational constraints.

5.2.2 Analytic model for reionization history

In this section, we present the set of analytic equations that tracks the reionization history of intergalactic hydrogen under the influence of the predicted galaxy populations. The model used in this work is similar to the ones presented in Madau et al. (1999, see also Choudhury 2009; Finkelstein et al. 2012a, 2019; Kuhlen & Faucher-Giguère 2012; Shull et al. 2012; Robertson et al. 2015; Madau 2017; Carucci & Corasaniti 2019; Naidu et al. 2020), modified to fully utilize the predictions from the Santa Cruz SAM for galaxy formation. With this model, we can efficiently predict volume-averaged ionizing photon emissivity ($\dot{n}_{\rm ion}$), IGM ionized fraction ($Q_{\rm HII}$), and the Thompson scattering optical depth ($\tau_{\rm CMB}$). In conjunction with the Santa Cruz SAM, the full modelling pipeline effectively connects the 'ground-level' galaxy formation physics to the 'top-level' cosmic reionization-related observables. With this modelling pipeline, we explore and test the impact of individual model components and how they impact the cosmological scale observables. Note that predictions for helium reionization are beyond the scope of this work.

Ionized volume fraction

The temporal evolution of the volume-averaged ionizing volume-filling fraction of ionized hydrogen, $Q_{\rm HII}$, is described by the first-order differential equation

$$\frac{dQ_{\rm HII}}{dt} = \frac{\dot{n}_{\rm ion}}{\bar{n}_{\rm H}} - \frac{Q_{\rm HII}}{\bar{t}_{\rm rec}}, \tag{5.2}$$

derived in Madau et al. (1999). The two terms can be interpreted as a growth term and a sink term, respectively, where the former is the ratio of the comoving ionizing emissivity, $\dot{n}_{\rm ion}$, and the volume-averaged comoving number density for intergalactic hydrogen, $\bar{n}_{\rm H}$; the latter is characterized by the ionized volume fraction divided by the recombination time-scale of ionized hydrogen, $\bar{t}_{\rm rec}$. We adopted $\bar{n}_{\rm H} = 1.9 \times 10^{-7}$ cm^{-3} as reported by Madau & Dickinson (2014).

Ionizing emissivity

The comoving emissivity of ionizing photons, $\dot{n}_{\rm ion}$, is the total budget supplied to reionize the IGM by galaxies, which is commonly modelled as the product of cosmic SFR or UV density, the LyC production efficiency of ionizing photons, and the fraction of photons that escapes to the IGM

$$\dot{n}_{\rm ion} = f_{\rm esc}\ \xi_{\rm ion}\ \rho_{\rm UV}. \tag{5.3}$$

Recalling that in our models, $f_{\rm esc}$ and $\xi_{\rm ion}$ may have a different value for each galaxy, instead of combining these for the whole population as above, we calculate the comoving value at each redshift by summing over all predicted galaxies

$$\dot{n}_{\rm ion} = \sum_i n_{{\rm h},i}\ \dot{N}_{{\rm ion},i}\ f_{{\rm esc},i}, \tag{5.4}$$

where $n_{\rm h}$ is the number density per Mpc^3 for each galaxy i, assigned based on the virial mass of the host halo (Section 5.2.1), $\dot{N}_{\rm ion}$ is the intrinsic ionizing photon production rate (Section 5.2.1), and $f_{\rm esc}$ is the LyC escape fraction (Section 5.2.1). This modified approach does not require a predetermined truncation value of $M_{\rm UV}$, as the turnover in the galaxy UV LF is a physical feature of our model. Moreover, Fig. 2.7 of Chapter 2 and Fig. 5.2 have shown that the magnitude where the UV LF turns over is directly correlated with the faint-end slope, which are both affected by the SN feedback slope $\alpha_{\rm rh}$. Therefore, instead of exploring a range of LF faint-end slope as is frequently done in other studies, we explore a range of $\alpha_{\rm rh}$.

Intergalactic HII recombination time-scale

The recombination time-scale for intergalactic hydrogen is given by

$$\bar{t}_{\rm rec} = \left[C_{\rm HII}\ \alpha_{\rm B}(T)\ (1+\eta Y/4X)\ \bar{n}_{\rm H}\ (1+z)^3\right]^{-1}, \tag{5.5}$$

where $C_{\rm HII}$ is a redshift-dependent HII clumping factor and $\eta = 1$ for singly ionized helium at $z > 4$. We adopted numerical predictions for the clumping factor from the radiation-hydrodynamical simulation L25N512 by Pawlik, Schaye, & Vecchia (2015), which $C_{\rm HII}$ evolves from ~ 1.5 to ~ 4.8 between $z \sim 14$ to ~ 6. The quantity $\alpha_B(T)$ is the temperature-dependent case B recombination coefficient for hydrogen given in Hui & Gnedin (1997), where we adopt $T = 2 \times 10^4$ K for the temperature of the IGM at the mean density; $\bar{n}_{\rm H}$ is the mean density of hydrogen in the IGM. Under the limitation of this type of model, we assume homogeneous recombination.

Thompson scattering optical depth of the CMB

The reionization history, $Q_{\rm HII}(z)$, is obtained by solving eqn. 5.2 using Python tools from `scipy.integrate.odeint` and `astropy.cosmology` (Robitaille et al. 2013; Price-Whelan et al. 2018; Virtanen et al. 2020). We can then calculate the Thomson scattering optical depth of the CMB, $\tau_{\rm CMB}$, using

$$\tau_{\rm CMB} = \int_0^\infty dz \frac{c(1+z)^2}{H(z)} Q_{\rm HII}(z) \sigma_{\rm T} \bar{n}_{\rm H} (1 + \eta Y/4X), \tag{5.6}$$

where $H(z)$ is the Hubble constant and $\sigma_{\rm T}$ is the Thomson cross section.

The main components of our default 'reference model', used throughout the remainder of this work are summarized in Table 5.1. This approach provides quick estimates of the volume-averaged reionization history and other cosmological-scale observables. However, it does not track the growth of individual Stromgren spheres. It also does not account for local density variances (e.g. void or over-dense regions), which may significantly affect the reionization histories on small scales. We will further discuss the limitations of the model in Section 5.4.

Table 5.1 Summary of components for reference model.

Model / Constraints	References	Configurations
Star formation	Bigiel et al. 2008	two-slope ($1 \to 2$)
Gas partitioning	Gnedin & Kravtsov 2011	metallicity-based
Stellar feedback	Somerville et al. 2008	$\alpha_{\rm rh} = 2.4$, $\varepsilon_{\rm SN} = 1.7$
LyC productivity	Stanway & Eldridge 2018	binary, v2.2.1
HII recombination	Hui & Gnedin 1997	Case B
HII clumping factor	Pawlik et al. 2015	L25N512 simulation
Emissivity constraints	Becker & Bolton 2013	$\dot{n}_{\rm ion}$ at $z = 2$–5
CMB constraints	Planck Collaboration 2016b	$\tau_{\rm CMB} = 0.058 \pm 0.012$

5.3 IGM Reionization by High-Redshift Galaxies

In this section, we present a collection of predicted reionization histories and investigate how galaxy formation physics can affect these predictions. We experiment with a range of constant values of $f_{\rm esc}$ (Section 5.3.1) or treat it as a function of redshift (Section 5.3.2). We also present a comparison with two other analogous studies (Section 5.3.3). At the end of the section, we probe the contribution of galaxies from different rest-$M_{\rm UV}$, as well as forecasting the contribution from galaxies observable by *JWST*.

5.3.1 Constant escape fraction

At first, we take the simplest approach by letting $f_{\rm esc}$ be a non-evolving, universal quantity. We present predictions for $\dot{n}_{\rm ion}$, $Q_{\rm HII}$, and $\tau_{\rm CMB}$ using our reference model configurations. Taking advantage of the efficiency of our modelling pipeline, we then performed a controlled experiment by varying a set of selected model components to quantity their impact on these predictions.

In Fig. 5.4, we show the evolution of $\dot{n}_{\rm ion}$ predicted by the reference model assuming $f_{\rm esc} = 0.20$. These results are compared to constraints on the global LyC emissivity at $2 < z < 5$ derived from the high-redshift Lyα forest by BB13. The plotted data are for the fiducial temperature-density parameter $\gamma = 1.4$ and spectral index of the ionizing sources $\alpha = 2.0$, with shaded area showing the reported total error. Historically, there has been tension between the Thomson scattering optical depth of the CMB, $\tau_{\rm CMB}$ and the ionizing photon emissivity at intermediate redshift, as discussed in the introduction. To demonstrate how these new constraints have eased the tension, we also show the compilation of constraints presented in Kuhlen & Faucher-Giguère (2012), which includes Lyα forest observations from Bolton & Haehnelt (2007); Faucher-Giguère et al. (2008b); Prochaska et al. (2009); Songaila & Cowie (2010). The BB13 constraints are a factor of ~ 2 higher than the previous measurements, and no longer require the total LyC emissivity to decrease so rapidly at $z \lesssim 6$. We also show the critical comoving ionizing emissivity, $\dot{n}_{\rm ion}^{\rm crit}$, or the minimum $\dot{n}_{\rm ion}$ that is required to keep the Universe ionized

$$\dot{n}_{\rm ion}^{\rm crit} = C_{\rm HII}\, \alpha_{\rm A}(T)\, (1 + \eta Y/4X)\, \bar{n}_{\rm H}^2\, (1+z)^3, \tag{5.7}$$

obtained by inverting the recombination time-scale given in eqn. 5.5. Here, the temperature-dependent case A recombination coefficient for hydrogen, $\alpha_{\rm A}(T)$, given by Hui & Gnedin (1997) is invoked because direct recombination from free to the ground bound state is more likely to occur in the optically thin IGM. On the other hand, case B is more suitable at prescribing regions near a source given that photons release by free-to-ground recombination is likely to reionize a nearby by hydrogen in these denser regions (see Faucher-Giguère et al. 2009 for in-depth discussion). The rest of the variables are consistent with the ones adopted in our calculation of the HII recombination time-scale.

We compare these results with alternative scenarios predicted with a range of $f_{\rm esc}$ and SN feedback slopes $\alpha_{\rm rh}$. As explored in previous chapter of the thesis, we found that SN feedback is the dominant process that regulates star formation in low-mass halos. Deviating from the fiducial value $\alpha_{\rm rh} = 2.8$, we found that the range $\alpha_{\rm rh} = 2.0$ to 3.6 yield a range of faint-end slopes that are still well within the current observational uncertainties at $z \gtrsim 6$. As shown in Fig. 2.7 of Chapter 2 and in

Fig. 5.2, low-mass galaxies are more abundant when feedback is weaker ($\alpha_{\rm rh} = 2.0$) and, conversely, less abundant when feedback is stronger ($\alpha_{\rm rh} = 3.6$). Here, we show the range of predicted $\dot{n}_{\rm ion}$ for galaxy populations predicted with these boundary cases and found that these yield results nearly ~ 1 dex apart. Similarly, we also experiment with a wide range of $f_{\rm esc} = 0.05$ to 0.80 to quantify its impact on the overall emissivity. From these results, we can already see that the LyC emissivity is more sensitive to the escape fraction than the faint-end slope of the UV LFs, for these variables within a physically meaningful range.

To explore the uncertainties in modelling $\dot{N}_{\rm ion}$, we added predictions with $\dot{N}_{\rm ion}$ from BC03, which is the least optimistic model explored in Chapter 4, and a scenario with constant $\xi_{\rm ion}$ adopting the expression from Kuhlen & Faucher-Giguère (2012)

$$\dot{N}_{\rm ion} = 2 \times 10^{25} {\rm s}^{-1} \left(\frac{L_{\rm UV}}{{\rm erg\ s^{-1} Hz^{-1}}} \right) \zeta_{\rm ion}, \tag{5.8}$$

where $\zeta_{\rm ion}$ is a free parameter that characterizes the hardness of the spectra. Here we assumed $\zeta_{\rm ion} = 1$ as in the fiducial model of KF12. The rest-UV magnitude is converted using $\log_{10}(L_{\rm UV}/({\rm erg\ s^{-1} Hz^{-1}})) = 0.4(51.63 - M_{\rm UV})$. This is equivalent to adopting a constant $\log \xi_{\rm ion} = 25.30$. We find that models with $\dot{N}_{\rm ion}$ computed self-consistently from the SPS models result in a shallower growth in $\dot{n}_{\rm ion}$ over time comparing to the model with a constant $\xi_{\rm ion}$. This is likely due to ageing and metal enrichment in the stellar populations in these galaxies, which naturally make the production of ionizing photons less efficient, although the number density of galaxies is growing. However, this effect is insufficient to reproduce the BB13 constraints as the flattening due to changes in $\xi_{\rm ion}$ is quite subtle, and overall $\dot{n}_{\rm ion}$ is still largely dominated by the fairly rapid growth of the overall number density of galaxies. On the other hand, results using $\dot{N}_{\rm ion}$ predicted by different SPS models seems to have evolved quite similarly over time with the expected factor of ~ 2 offset due to the inclusion of binary stars in the BPASS models. For further discussion of differences between the BPASS binary SPS model and BC03, we refer the reader to the discussion associated with Fig. 4.12 in Chapter 4.

In the same spirit as the $\dot{n}_{\rm ion}$ comparison, in Fig. 5.5 we present the predicted IGM neutral fraction, $Q_{\rm HI} \equiv 1 - Q_{\rm HII}$, from the same set of model variants. These predictions are stacked up against a compilation of observational constraints compiled from R13 and R15, which consist of various kinds of observations, including Lyα emitting galaxies (Ota et al. 2008; Ouchi et al. 2010; Pentericci et al. 2014; Schenker et al. 2014), Lyα emission fraction (McQuinn et al. 2007; Mesinger & Furlanetto 2008; Dijkstra et al. 2011), Lyα galaxy clustering (Ouchi et al. 2010), Lyα damping wing (Totani et al. 2006; McQuinn et al. 2008; Chornock et al. 2013), from the near zones of bright quasars (Bolton & Haehnelt 2007; Bolton et al. 2011; Schroeder et al. 2013), and from dark pixels

in Lyα forest measurements (Mesinger 2010; McGreer et al. 2011, 2015). We refer the reader to Robertson et al. (2013) for a detailed description of these constraints. We also added the latest constraints from Lyα emitting galaxies reported by Mason et al. (2018b, 2019).

Fig. 5.6 shows $\tau_{\rm CMB}$ as a function of redshift for our reference model, and for the model variants $f_{\rm esc} = [0.05, 0.80]$ and $\alpha_{\rm rh} = [2.0, 3.6]$. We show recent measurements reported by the Planck Collaboration 2014; 2016b; 2018 and WMAP-9 (Hinshaw et al. 2013). The latest observational constraints together favour both a later conclusion of reionization and a less rapidly evolving $\dot{n}_{\rm ion}$, which ease both the need for high emissivity at high redshifts and rapid decrease of $\dot{n}_{\rm ion}$ toward $z \lesssim 5$. In Fig. 5.7, we show the integrated $\tau_{\rm CMB}$ as a function of both $f_{\rm esc}$ and $\alpha_{\rm rh}$. This shows that $\tau_{\rm CMB}$ is very sensitive to the LyC escape fraction for $f_{\rm esc} \lesssim 0.3$, but its dependence on $f_{\rm esc}$ becomes much flatter above this value. For $f_{\rm esc} \gtrsim 0.3$, the predicted optical depth is more sensitive to the abundance of faint galaxies rather than the LyC $f_{\rm esc}$. Note that $\tau_{\rm CMB}$ is an integrated quantity that compresses the reionization history into a single metric. However, it is degenerately affected by both the conclusion of the phase transition and its progression. For instance, an extremely slow reionization progression or a rapid, late reionization can both result in a lower measured $\tau_{\rm CMB}$ value.

These results show that models with non-evolving $f_{\rm esc}$ and $\alpha_{\rm rh}$ are unable to reproduce a reionization history that simultaneously matches all three sets of constraints, which are consistent with results from previous studies (e.g. Kuhlen & Faucher-Giguère 2012; Anderson et al. 2017).

5.3.2 Constraining a redshift dependent escape fraction with MCMC

Results from §5.3.1 quantified the sensitivity of model outputs to a range of fixed values of $f_{\rm esc}$ and $\alpha_{\rm rh}$. In this section, we allow $f_{\rm esc}$ to evolve as a function of redshift (see eqn.5.1) and employed a Markov chain Monte Carlo (MCMC) method to find the optimal configuration that is needed to satisfy the current observational constraints. We employ the python MCMC tool `emcee`[3] by Foreman-Mackey et al. (2013) to survey the four dimensional parameter space, including $f_{\rm esc,max}$, $f_{\rm esc,0}$, k_0, and $\alpha_{\rm rh}$. In the context of cosmic reionization studied here, we consider the many other free parameters in the galaxy formation model as being collectively constrained either by calibration or by the cross-checks with observations between $z = 4$ to 10 in previous works. In this exercise, $\alpha_{\rm rh}$ can take any value within the range [2.0, 3.6], where $\dot{n}_{\rm ion}$ is pre-calculated for fixed values of $\alpha_{\rm rh} = 2.0$, 2.4, 2.8, 3.2, and 3.6, and then interpolated using the `scipy.interpolate` tool. By varying $\alpha_{\rm rh}$ within the SAM, we have included a number of associated features under its influences, including the faint-end slopes and flattening of the UV LFs, and the subtle boost of ionizing photon

[3] http://dfm.io/emcee, v2.2.1

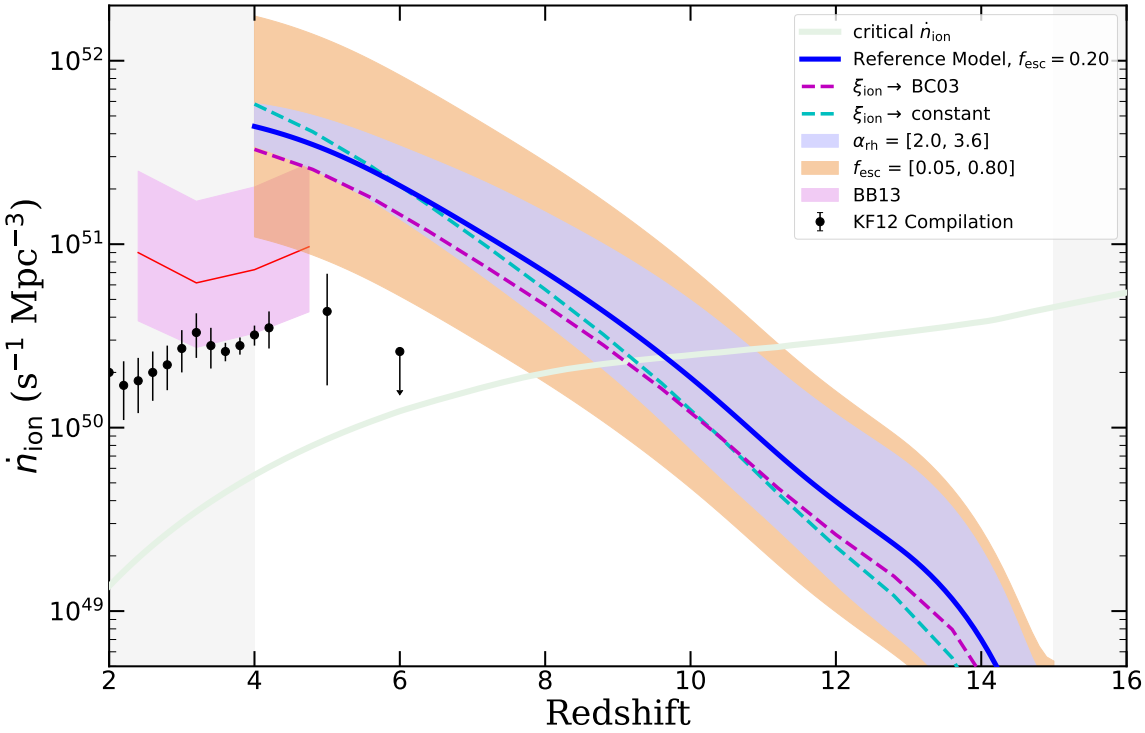

Figure 5.4 Ionizing photon emissivity, $\dot{n}_{\rm ion}$, as a function of redshift predicted with the reference model configurations (blue solid, see Table 5.1). The blue shaded region marks the range of $\dot{n}_{\rm ion}$ from galaxy populations predicted for $\alpha_{\rm rh} = 2.0$ (weaker feedback, leading to a higher number density of low-mass galaxies and higher emissivity) and 3.6 (stronger feedback, leading to fewer low-mass galaxies and lower emissivity). The orange region marks the range predicted with $f_{\rm esc} = 0.05$ (lower emissivity due to low escape fraction) and 0.80 (higher emissivity). We also include predictions made with a constant $\log(\xi_{\rm ion}) = 25.30$ and with $\dot{N}_{\rm ion}$ from the SPS models of BC03. These results are compared to observational constraints from BB13 and a compilation from KF12. The light green band shows the critical ionizing photon emissivity required to keep the Universe ionized (see eqn. 5.5 and associated description in text). This shows how uncertainties in different model components could have affected the total ionizing budget throughout the EoR.

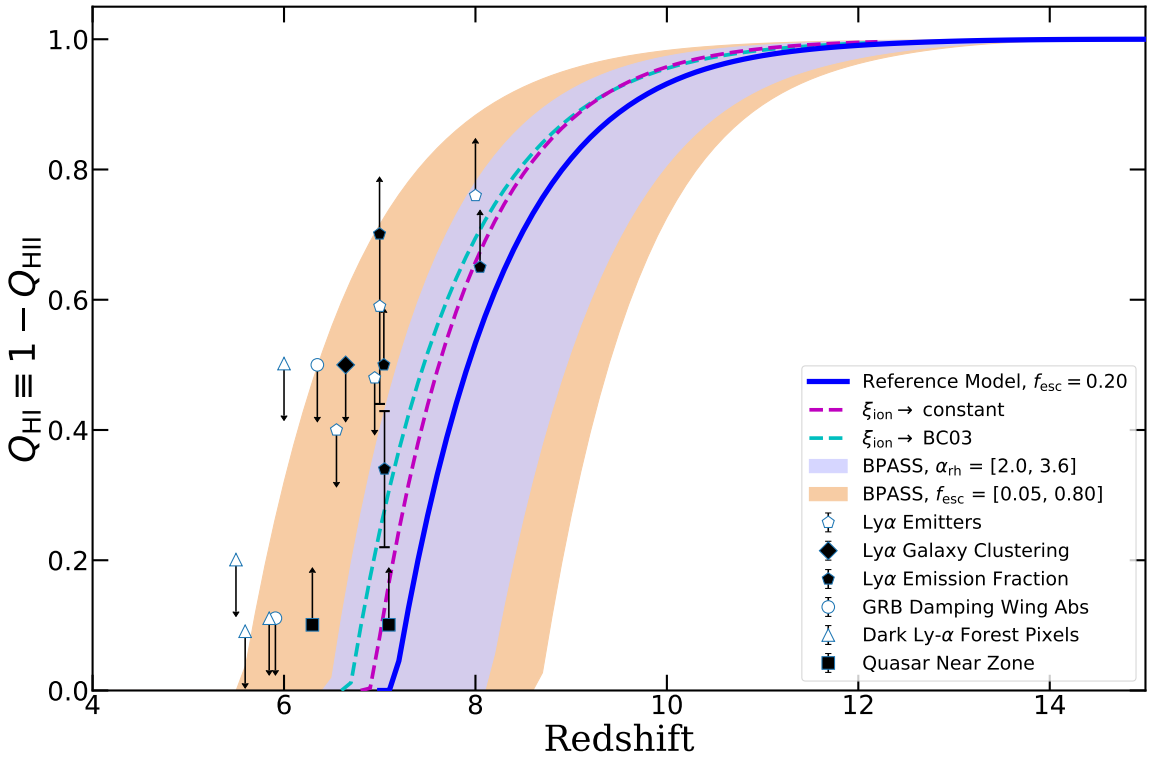

Figure 5.5 Neutral volume fraction, $Q_{\rm HI} \equiv 1 - Q_{\rm HII}$, as a function of redshift calculated for our reference model (blue solid). The blue shaded region marks the range for galaxy populations predicted for $\alpha_{\rm rh} = 2.0$ (weaker feedback; lower bound) and 3.6 (stronger feedback; upper bound). The orange region marks the range predicted with $f_{\rm esc} = 0.05$ (upper bound) and 0.80 (lower bound). We also include predictions made with a constant $\log(\xi_{\rm ion}) = 25.30$ and with $\dot{N}_{\rm ion}$ from BC03. These results are compared to a compilation of observational constraints from R13 and R15, and additional constraints from Mason et al. (2018b, 2019). The simple reference model, with fixed $f_{\rm esc}$, is in tension with these observations, though it is clear that this is primarily due to uncertainties on $f_{\rm esc}$.

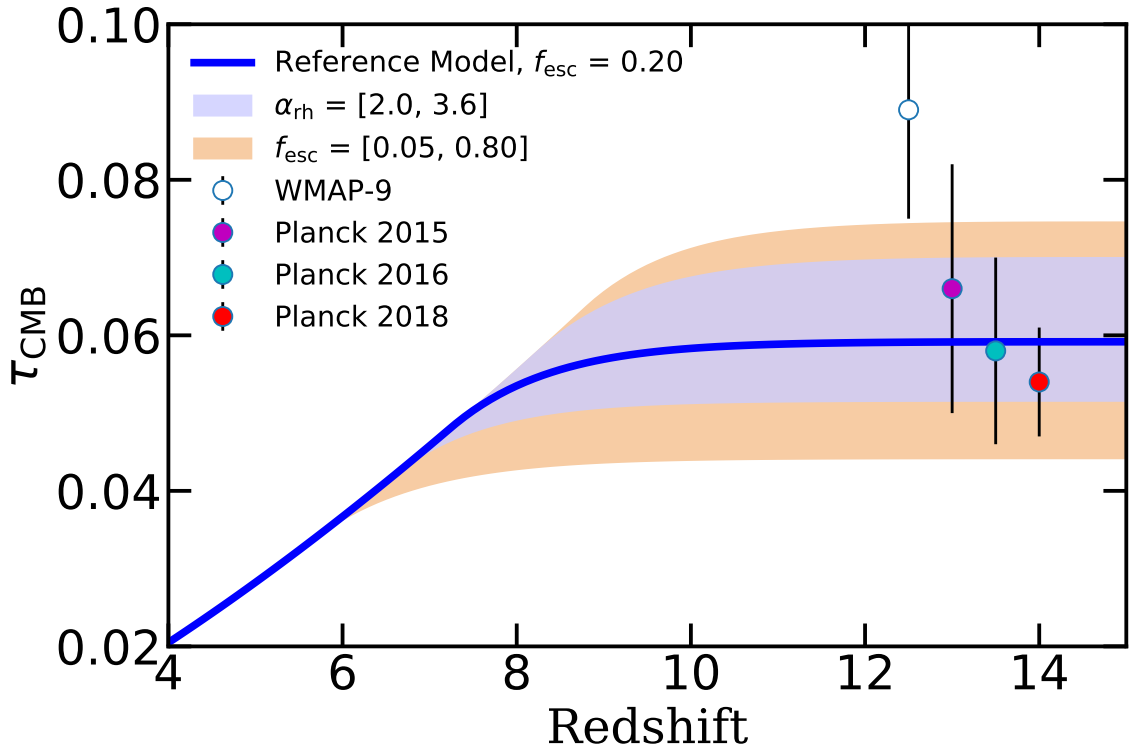

Figure 5.6 Thomson scattering optical depth of the CMB, $\tau_{\rm CMB}$, as a function of redshift predicted by our reference model (blue solid). The blue shaded region marks the range of integrated $\tau_{\rm CMB}$ from galaxy populations predicted for $\alpha_{\rm rh} = 2.0$ (upper bound) and 3.6 (lower bound). The orange region marks the range predicted with $f_{\rm esc} = 0.05$ (lower bound) and 0.80 (upper bound). We also show reported values from the Planck Collaboration 2014; 2016b; 2018 and from WMAP-9 (Hinshaw et al. 2013). Our reference model has no difficulty accounting for the more recent estimates of $\tau_{\rm CMB}$.

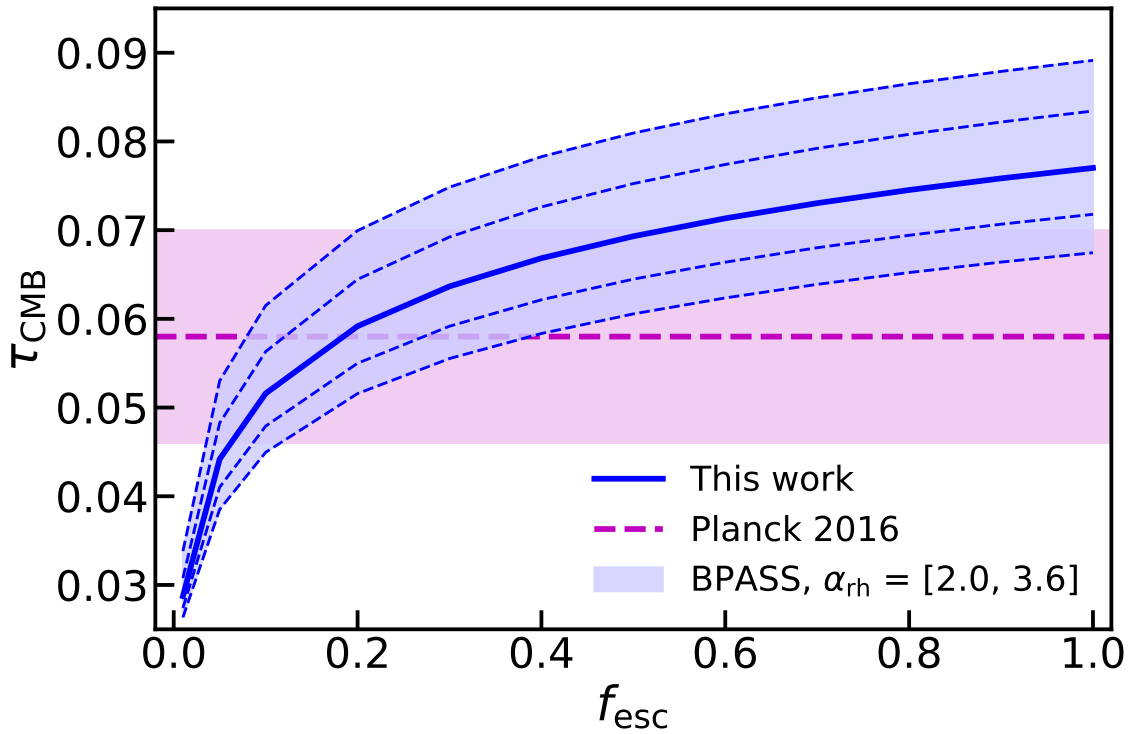

Figure 5.7 Thomson scattering optical depth of the CMB, $\tau_{\rm CMB}$, as a function of assumed $f_{\rm esc}$ and $\alpha_{\rm rh}$ in our model. Predictions from our reference model are shown by the blue solid line. The blue dashed lines show the alternative predictions made with $\alpha_{\rm rh} = 2.0$ (upper bound for earlier reionization due to higher number density of galaxies), 2.4, 3.2, and 3.6 (lower bound for later reionization due to lower number density of galaxies). The measured value of $\tau_{\rm CMB} = 0.058 \pm 0.012$ as reported by Planck Collaboration (2016b) is shown for comparison. This provides a reference showing the interplay between uncertainties or variation in the parameters $f_{\rm esc}$ and $\alpha_{\rm rh} = 2.0$.

production due to the slight increase in burstiness triggered by SN feedback (see Section 4.3.4 and associated discussion in Chapter 4).

The MCMC framework is set up with 56 walkers, each of which performs a chain of 120,000 steps with the first 200 regarded as burn-in and discarded. Each of these walkers is initialized with a Gaussian distribution, with a chosen peak and half-width distribution. The parameters that went into the set-up can be found in Table 5.2. We assumed a flat prior for all four of our free parameters. For a randomly drawn prior that falls outside the boundary of the flat prior, a new set of parameters are drawn.

The set of observational constraints used in the MCMC are the Lyα forest constraints on $\dot{n}_{\rm ion}$ from BB13 and the $\tau_{\rm CMB}$ from Planck Collaboration (2016b), which are weighted equally in the likelihood function. Note that the large collection of IGM neutral fraction estimates are shown for comparison but are not used as constraints in the MCMC. The median and the 68% and 95% confidence region of our posteriors for the predicted $f_{\rm esc}$, $\dot{n}_{\rm ion}$, $Q_{\rm HI}$, and $\tau_{\rm CMB}$ are summarized in Fig. 5.8, where the posterior distributions are shown in Fig. 5.9.

Our results favour a drop in escape fraction at $z \lesssim 7$, leading to a turnover in the ionizing emissivity. The parameters k_0 and $\alpha_{\rm rh}$ are strongly covariant, and are only weakly constrained by $\tau_{\rm CMB}$. It is encouraging that the range of predicted reionization histories are in broad agreement with the $Q_{\rm HI}$ constraints. This is non-trivial as it depends on other model components that are not being actively 'tuned' here, such as which galaxies are contributing to reionization. We note that the latest Planck Collaboration 2018 results would favour an even milder evolution of $f_{\rm esc}$ and a slightly lower $f_{\rm esc,max}$. Adopting the Planck2018 value of $\tau_{\rm CMB}$ will only mildly change the results and conclusions of this work.

Table 5.2 Summary for the MCMC parameters, flat prior constraints, and posterior with 68% confidence region.

	Initiation	σ	constraints	posterior
$f_{\rm esc,0}$	0.036	0.00005	[0.012, 0.060]	$0.0381^{+0.0148}_{-0.0159}$
$f_{\rm esc,max}$	0.350	0.0005	[0.100, 0.500]	$0.2985^{+0.1357}_{-0.1328}$
k_0	0.50	0.005	[0.10, 0.90]	$0.523^{+0.255}_{-0.269}$
$\alpha_{\rm rh}$	2.80	0.05	[2.0, 3.6]	$2.784^{+0.525}_{-0.511}$

5.3.3 Comparison with other recent models

In this section, we compare the predicted reionization history in our reference model to recent studies by Finkelstein et al. (2019) and Naidu et al. (2020). In Fig. 5.10, we compare results from

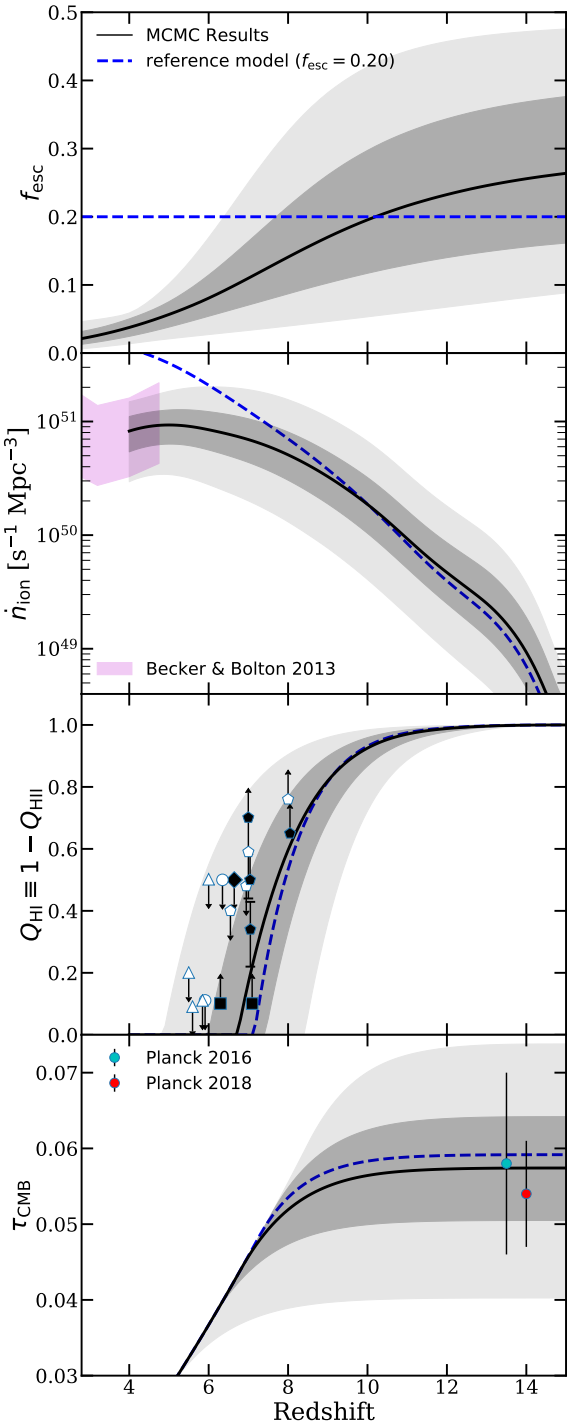

Figure 5.8 From top to bottom, we show the predicted redshift evolution for $f_{\rm esc}$, $\dot{n}_{\rm ion}$, $Q_{\rm HI}$, and $\tau_{\rm CMB}$ using the BB13 and Planck Collaboration (2016b) $\tau_{\rm CMB}$ observations as constraints in our MCMC analysis. The shaded areas denote the 68% (dark grey) and 95% (light grey) confidence regions. We also show our reference model with a constant $f_{\rm esc} = 0.20$ for comparison (blue dashed). This figure shows that a moderate effective evolution of $f_{\rm esc}$ with redshift can comfortably accommodate all of the observational constraints.

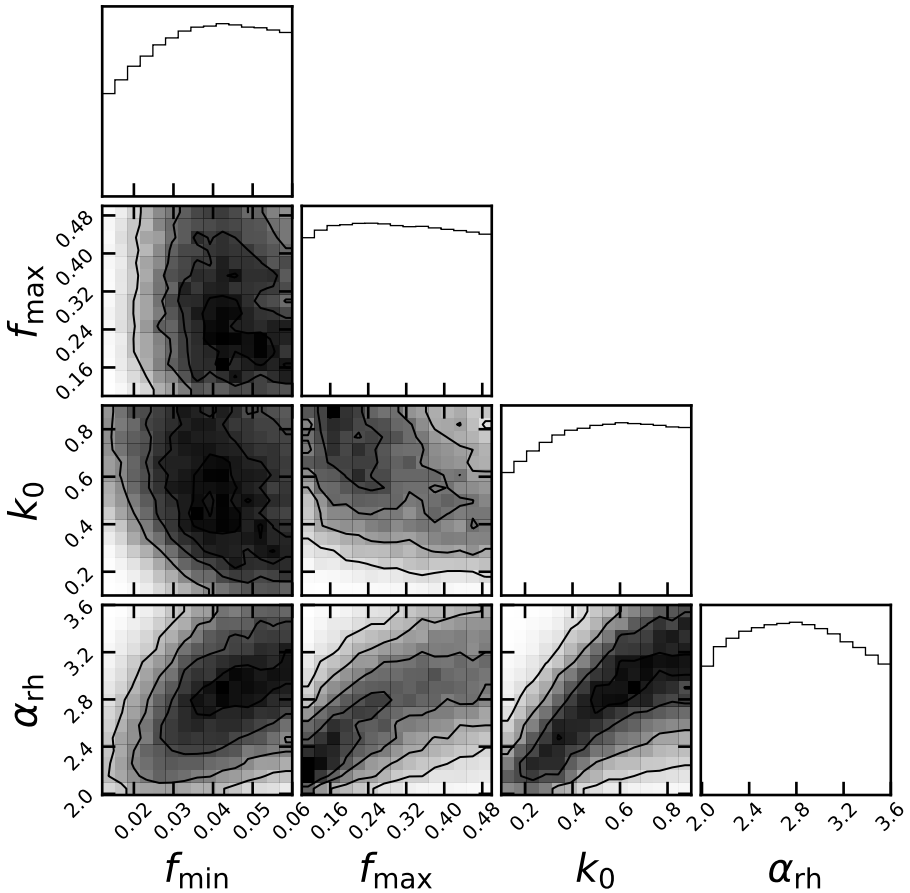

Figure 5.9 Distributions of the parameter posterior distributions from the MCMC. The parameter $\alpha_{\rm rh}$ is fairly well constrained, as is the low-redshift (here $z = 4$) value of $f_{\rm esc}$ ($f_{\rm esc,0}$). The rate of evolution of $f_{\rm esc}$ (k_0) and its asymptotic value ($f_{\rm esc,max}$) are not well constrained. The strong covariance between $\alpha_{\rm rh}$ and $f_{\rm esc}$ is apparent.

our reference model with a constant $f_{\rm esc} = 20\%$ and with the evolving $f_{\rm esc}$ found in Section 5.3.2 to the results from Finkelstein et al. and Model I (constant $f_{\rm esc}$) from Naidu et al.. Although these models are fairly similar as they adopted a similar approach to modelling reionization, we note that these works adopted very different approaches to model galaxy populations and their evolution.

Finkelstein et al. use UV LFs from observations with faint-end slopes extrapolated below the current detection limit. These UV LFs are truncated at halo masses corresponding to photoionization squelching and atomic cooling, which are obtained via abundance matching. The many moving parts in the model, including the LyC escape fraction, halo truncation mass, the evolution of $\xi_{\rm ion}$, and the contribution of AGN, are optimized to fit a set of observational constraints using an MCMC machinery. Finkelstein et al. were particularly interested in exploring models that could satisfy all constraints on reionization while adopting a low ionizing escape fraction ($\lesssim 5\%$ throughout the EoR). They adopted a halo-mass dependent parametrization of $f_{\rm esc}$ based on hydrodynamic simulations from Paardekooper et al. (2015), and parametrized $\xi_{\rm ion}$ in terms of redshift and galaxy luminosity.

On the other hand, Naidu et al. adopted galaxies from the Tacchella et al. (2018) empirical model (see comparison with our predictions in Chapter 3) and estimated the production efficiency of ionizing photons using synthetic SEDs generated from the Flexible Stellar Population Synthesis (FSPS; Conroy et al. 2009, 2010) and MESA Isochrones and Stellar Tracks (MIST; Choi et al. 2017b) for individual galaxies. They have explored a model with fixed $f_{\rm esc}$ and one that scales as a function of $\Sigma_{\rm SFR}$, and a range of truncation values $M_{\rm UV}$.

Finkelstein et al. found that in order for models with such universally low escape fractions to be viable, a rather high and rapidly evolving $\xi_{\rm ion}$ is required. The range of $\xi_{\rm ion}$ values are similar to, or even above, the observed values from Bouwens et al. (2016c). However, Chapter 4, Wilkins et al. (2016b), and Ceverino et al. (2019) have shown that such high values of $\xi_{\rm ion}$ and such strong evolution are not 'naturally' predicted in current self-consistent galaxy formation models. As we can see in Fig. 5.10, the Finkelstein et al. model (in which reionization is heavily dominated by low-mass galaxies) predicts an early start to reionization and a more gradual evolution for $Q_{\rm HI}$. The Naidu et al. model (in which massive galaxies play a more important role) predicts later and more rapid reionization. Curiously, our model lies somewhere in between, although it also predicts a fairly rapid transition in $Q_{\rm HI}$.

This comparison illustrates that there is still significant uncertainty in which galaxies dominate the reionization of the Universe and the details of how reionization progressed. Future observations *JWST* and other facilities will provide direct constraints on the source populations (as we explore further in the next section). Furthermore, as galaxies of different masses cluster very differently in space, these models would also have very different implications for the topology of reionization,

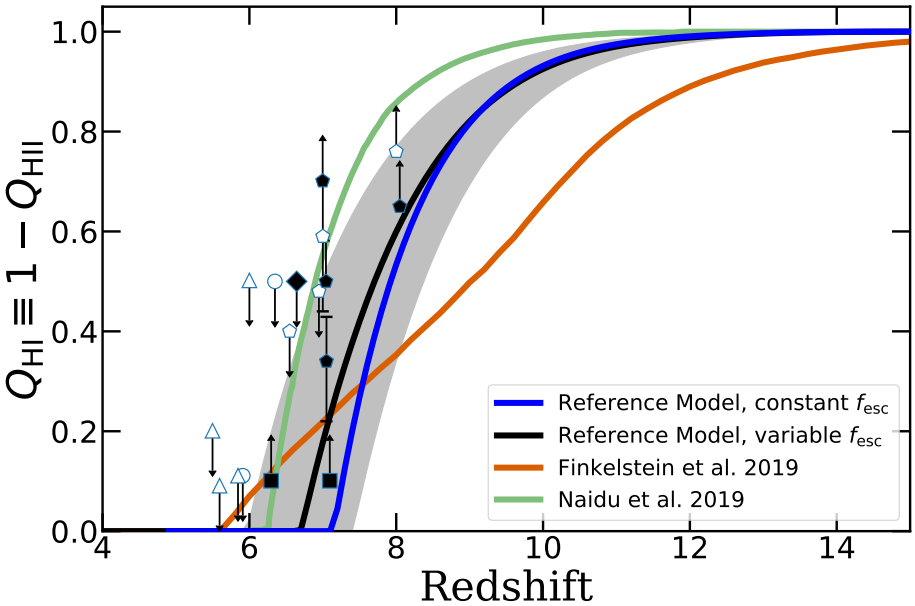

Figure 5.10 The redshift evolution of the IGM neutral fraction predicted by our reference model with a constant 20% (blue) escape fraction, and our fiducial model with an evolving (black line and shaded regions) escape fraction, compared to predictions from the recent models of Finkelstein et al. (2019) (orange) and Naidu et al. (2020) (green). A compilation of observational constraints is shown by the symbols (see Fig. 5.5 for a legend). This illustrates the broad range of reionization histories implied by several of the most recent modelling papers.

which will eventually be probed with 21-cm intensity mapping experiments.

5.3.4 Which galaxies reionized the Universe and will JWST see them?

In this section, we take advantage of the completeness of our predictions both in mass and redshift to estimate the contributions of ionizing photons from galaxies of different intrinsic luminosities, and estimate what fraction of the ionizing photon budget will be contributed by galaxies that are anticipated to be observed in future *JWST* surveys. We show predictions for our reference model. As the dependence of $f_{\rm esc}$ on galaxy luminosity is very uncertain, and not yet included in our modelling, we only provide predictions here for the fraction of ionizing photons *produced* and do not try to estimate the fraction that escapes to the IGM. Recent simulation works have shown that $f_{\rm esc}$ may inversely scale with $M_{\rm h}$ in a fairly loose way (e.g. Paardekooper et al. 2015), and therefore the actual contribution from massive/luminous galaxies to the overall ionizing photon budget might be smaller than what is presented here. However, our models do incorporate a mass and redshift dependent $\xi_{\rm ion}$ based on our self-consistent modelling. These calculations do not account for field-to-field variance nor the survey area, where rare, massive objects may be missing from the small survey area of deep surveys.

In fig. 5.11, we show the fraction of $\dot{n}_{\rm ion}$ produced by galaxies above the detection limits of hypothetical *JWST* wide, deep, and lensed surveys with detection limits of $m_{\rm F200W} = 28.6$, 31.5, and 34.0, respectively[4]. See Table 2.6 in Chapter 2 and Table 3.1 in Chapter 3 for detailed configurations of these hypothetical surveys. For comparison, we also show results for legacy *HST* surveys, where we adopted detection limits for the F160W filter $m_{\rm F160W} = 26.8$, 29.5, and 31.5 for wide, deep, and lensed surveys, respectively, with configurations similar to the CANDELS and Hubble Frontier Fields surveys. At $z \sim 7$–8, where we predict the Universe to be about 50% reionized by volume, *JWST* will be able to detect the sources of 60-70% of the reionizing photons in a deep survey. This fraction increases to $\sim 90\%$ for an ultra-deep lensed survey, however, interpreting lensed observations and estimating the survey completeness may be more challenging.

In a similar experiment, we break down the galaxy populations by rest-frame *intrinsic* UV magnitude (not accounting for the effect of dust attenuation) into the following groups: $-24 < M_{\rm UV} < -20$, $-20 < M_{\rm UV} < -16$, $-16 < M_{\rm UV} < -12$, to the faintest $-12 < M_{\rm UV} < -8$. In fig. 5.12, we compare the fraction of $\dot{n}_{\rm ion}$ contributed by galaxies from each of these groups from $z = 15$ to 4. Galaxies beyond this range combined produce $< 1\%$ of ionizing photons across all redshifts, and are omitted here. Similar to results presented in the previous figure, we assume that $f_{\rm esc}$ does not depend on galaxy properties, which may significantly effect the predictions shown here. We find that ultra-faint galaxies ($-12 < M_{\rm UV} < -8$) dominate at the highest redshifts ($z \gtrsim 13$), with a slightly brighter population $-16 < M_{\rm UV} < -12$ dominating over the redshift range $10 \lesssim z \lesssim 13$. At lower redshift $z \lesssim 10$, galaxies in the intermediate luminosity range $-20 < M_{\rm UV} < -16$ dominate.

Similarly, in Fig. 5.13 we break down the contribution of ionizing photons by the host halo masses of galaxies. These are based on the predictions from our reference model configurations, and are quite sensitive to the details of how galaxies populate halos, which as we have shown depends on the details of the stellar feedback parameters and other physical processes. We find that contributions from halos outside the range shown here are insignificant. This result is also useful for estimating the 'completeness' of the predicted ionizing emissivity from studies with limited mass resolution. In Chapter 2 (see Section 2.2.2 and Fig. 2.3), we explored the impact on star formation from a photoionizing background using a redshift-dependent characteristic mass approach as described by Okamoto et al. (2008) and found nearly no impact on the galaxy populations at the range of redshift and halo mass relevant to our study. However, high-resolution hydrodynamic simulations have shown that the presence of such a background may have affected the low-mass, 'photosensitive' halos of $\log(M_{\rm h}) \lesssim 9$ (Finlator et al. 2013). Accounting for this effect may reduce the contribution from low-mass halos near the beginning of the EoR relative to our predictions.

[4]The F200W filter on the NIRCAM instrument

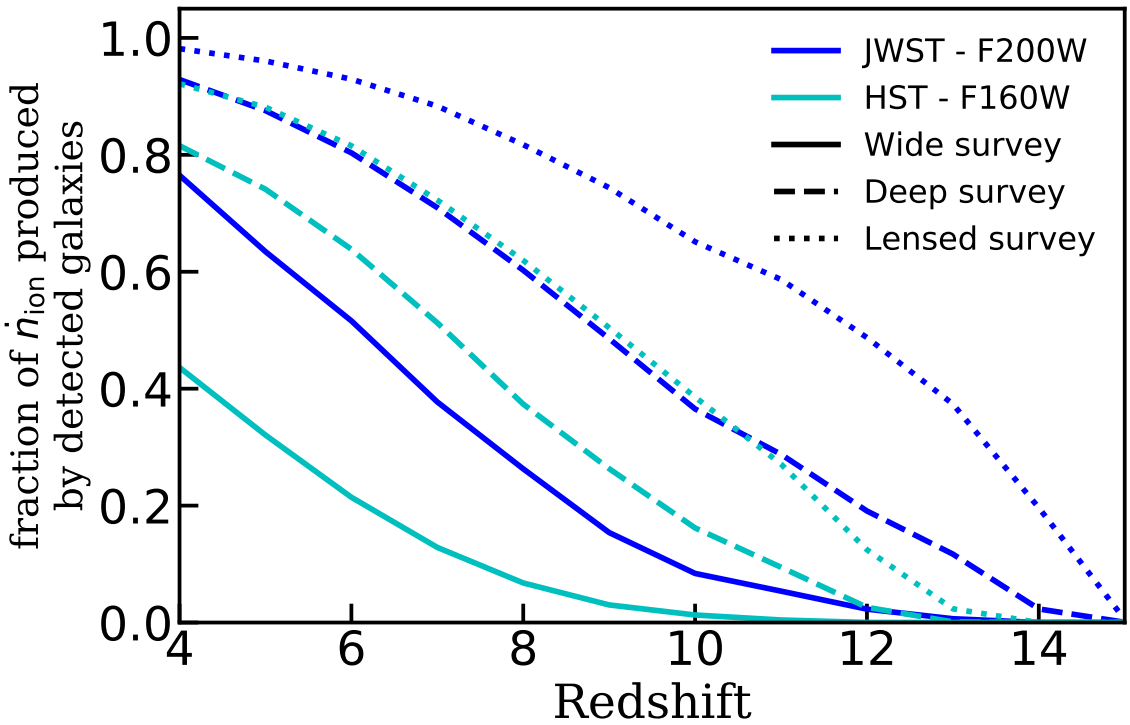

Figure 5.11 Predictions from our reference model for the fraction of ionizing photons produced by galaxies that are expected to be detected in various types of surveys, including wide (solid), deep (dashed), and lensed (dotted) surveys with *JWST* (blue) and *HST* (cyan) between $z = 4$ and 15. Survey areas and detection limits assumed for these calculations are detailed in the text. These predictions reflect the production rate and do not account for the escape fraction of ionizing photons possibly varying across galaxies with different luminosities.

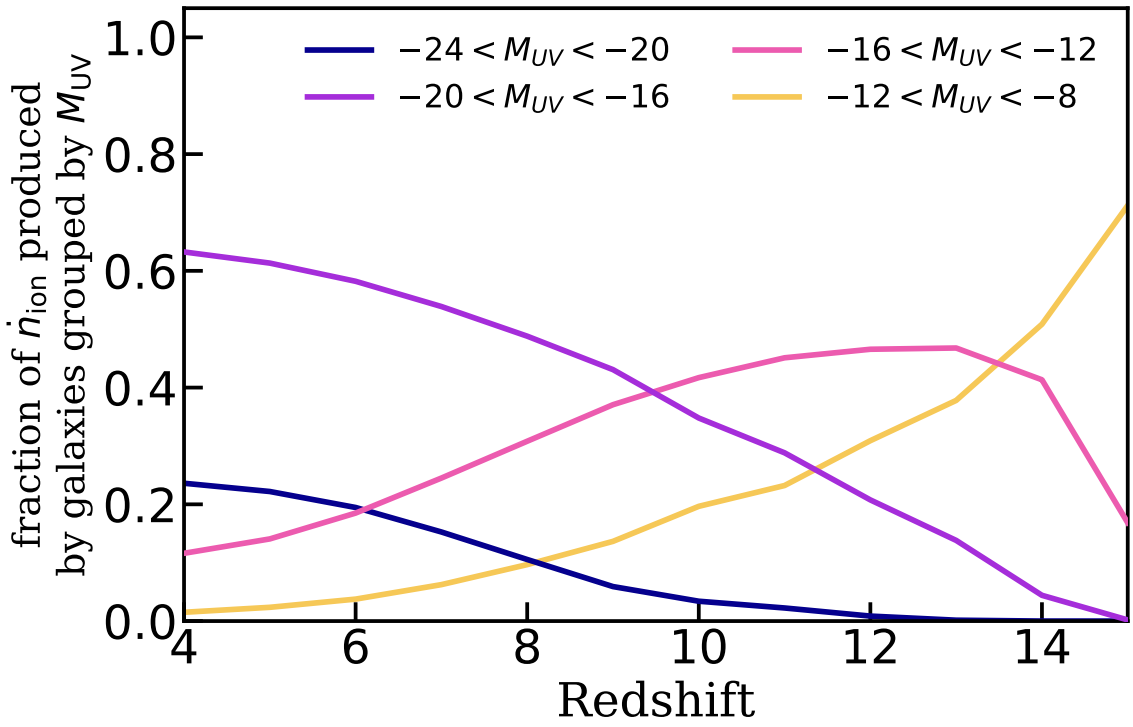

Figure 5.12 Predictions from our reference model for the fraction of ionizing photons produced by galaxies grouped by rest-frame dust-attenuated $M_{\rm UV}$ between $z = 4$ and 15. These predictions reflect the production rate and do not account for the escape fraction of ionizing photons possibly varying across galaxies with different luminosities. Galaxies outside the range of $M_{\rm UV}$ shown contribute $< 1\%$ of ionizing photons at all times.

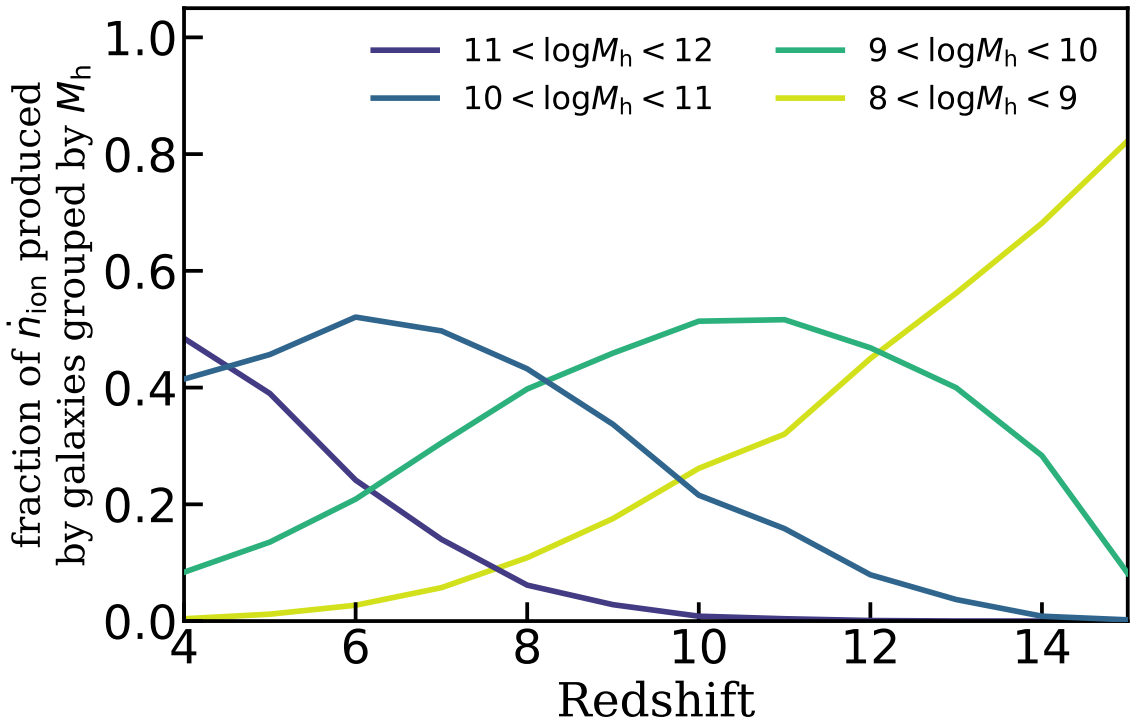

Figure 5.13 Predictions from our reference model for the fraction of ionizing photons produced by galaxies grouped by host halo mass $M_{\rm h}$ between $z = 4$ and 15. These predictions reflect the production rate and do not account for the escape fraction of ionizing photons possibly varying across halo mass. The contribution of ionizing photons that originates in halos outside the range shown is insignificant.

5.4 Discussion

In this section, we discuss some caveats and uncertainties in our modelling pipeline, and present an outlook for future observations with *JWST* and beyond.

5.4.1 Galaxies forming at extreme redshifts and their role in cosmological events

In this thesis, we have explored the interplay between galaxy formation physics and the cosmological-scale phase transition of hydrogen reionization. In particular, we investigated whether models with physical recipes and parameters that have been calibrated to match lower redshift observations ($z \sim 0$) are consistent with a broad suite of observations at extremely high redshifts ($z \gtrsim 6$). A significant finding of this work is that *these locally calibrated models are consistent (within the uncertainties) with all currently available observations at $z \gtrsim 6$, including direct observations of galaxies, and indirect probes of the reionization history from observations of the IGM and CMB.* This has two important implications: 1) It seems that the physical processes regulating star formation and stellar feedback do not operate in a vastly different manner at extremely high redshift. Given our lack of detailed understanding of how these processes work even in the local Universe, this is far from a trivial conclusion. 2) Contrary to some previous suggestions in the literature, the current suite of observations do not *require* an additional 'exotic' population of reionizing sources (other than galaxies, e.g. mini-quasars, Pop III stars, self-annihilating dark matter, etc.) in the early Universe. The remaining uncertainties on several components of our modelling mean that we do not *rule out* the existence of such sources at some level – but they are not required to satisfy existing constraints.

The overall ionizing photon budget available during the EoR is degenerately affected by physical processes that operate over a vast range of scales. As illustrated in Fig. 5.4, each of these seemingly degenerate components can evolve differently and be constrained independently. In Chapter 2 and Chapter 3, we provided physically motivated predictions for the evolution of the number density of galaxies at high redshift, which will be further constrained with future galaxy surveys; and in Chapter 4, we predicted the evolution and distribution of $\xi_{\rm ion}$ for the same set of galaxies, which may also be constrained with future observations as discussed in Chapter 4. We have explicitly broken down the contribution to the ionizing photon budget as a function of redshift from galaxies with different observed frame and rest-frame luminosity, and different halo mass. This is again a non-trivial calculation, as the intrinsic production efficiency of ionizing photons depends on a combination of factors such as stellar population age and metallicity in addition to the number density of galaxies with different luminosities. These effects are self-consistently included in our models.

We took a fully semi-analytic approach to assemble our modelling pipeline, including the construction of merger trees, formation and evolution of galaxies, and the progression of cosmic reionization. In practice, this modelling pipeline serves as a low-computational-cost platform for examining galaxy formation and cosmic reionization constraints from various tracers. As shown in Fig. 5.8, the set of IGM neutral fraction constraints seem to collectively favour a relatively rapid decline of neutral hydrogen around $z \sim 6$–7. However, such a reionization history would yield a $\tau_{\rm CMB}$ near the lower bound of the reported uncertainties of the latest measurements. Although these constraints are in mild tension, our model is in agreement with all constraints within the 68% confidence regions of the MCMC posterior. Adopting the even lower $\tau_{\rm CMB}$ measurement reported by the Planck Collaboration in 2018 would further ease this tension and yield a slightly milder evolution for $f_{\rm esc}$ and a slightly more gradual reionization history.

Another novel aspect of our work is the rigorous statistical exploration of the degeneracy in the physical parameter controlling the impact of stellar feedback on low-mass halos ($\alpha_{\rm rh}$) and the parametrized effective redshift evolution of $f_{\rm esc}$. Larger values of $\alpha_{\rm rh}$ result in stronger feedback, producing fewer low-mass galaxies, and require higher values of $f_{\rm esc}$ to produce the required budget of ionizing photons, and vice versa. Under the subset of observational constraints included in the MCMC, it is encouraging that the median of the posterior of $\alpha_{\rm rh}$ is very much in agreement with the value that is required to reproduce $z \sim 0$ observations. Similarly, the required redshift evolution in $f_{\rm esc}$ is not extreme, and the values at the lower end of our explored redshift range ($z \sim 4$) where there are some observational constraints are reasonable.

Past studies predicted diverging scenarios for the final stage of reionization depending on the assumptions and observational constraints employed by these models. Such process could be rather extended when dominated by low-mass galaxies (e.g. Finkelstein et al. 2019), or conversely very rapid when dominated by massive galaxies (e.g. Mason et al. 2018b and Naidu et al. 2020), for which the rapid end to reionization is motivated by Lyα emitter constraints. The results presented in this work depict a relatively early onset of reionization compared to Naidu et al. due to the early contributions from low-mass galaxies, but lag behind Finkelstein et al. because of the lower predicted $\xi_{\rm ion}$ and $f_{\rm esc}$. However, it is very intriguing to see that our model also predicted a very rapid end to reionization when Lyα emitters are not explicitly used to constrain our model. Fig. 5.12 shows that the contribution of ionizing photons from more massive galaxies has grown rapidly and took over from their low-mass counterparts during the EoR, which provides a physical explanation to the rapid conclusion of reionization that is solely driven by galaxy formation physics rather than from observed EoR constraints.

However, Finkelstein et al. also showed that by letting all galaxies to have the same escape

fraction, galaxies with $M_{\rm UV} \lesssim -16$ would have dominated the ionizing photon budget. Given that the way $f_{\rm esc}$ is parametrized in this work, it is also possible that we have overestimated the contribution from massive galaxies, which could be a partial reason to the rapid end. Therefore, the predicted rapid end to reionization can be one part backtracked to the predicted evolution of galaxy populations and their spectroscopic properties, and one part due to our parametrization of $f_{\rm esc}$.

5.4.2 Caveats, limitations, and uncertainties of the modelling framework

The limitations and caveats regarding the galaxy formation model and the physical recipe for $\dot{N}_{\rm ion}$ have been thoroughly discussed in previous works; we refer the reader to Section 3.6.3 in Chapter 3 and Section 4.4.3 in Chapter 4. This discussion will be focused mainly on the topics related to the reionization pipeline presented in this work.

We note that even though the SAM is fairly successful at reproducing a wide variety of existing observational constraints, which we examined in detail in Chapter 2 and Chapter 3, both the physical properties and number density of the predicted galaxy populations at $z > 10$ are poorly constrained due to the lack of direct observations. They are subject to uncertainties in model components, such as feedback effects and SF relations, which are either untested or known to be inaccurate in extreme (e.g. metal-free) environments. There are also missing physical processes, such as the formation of Population III stars, that can potentially affect star formation activity in low-mass halos in the early universe.

Therefore, we regard the predictions for $z = 11$ to 15, including both the UV LFs and $\dot{N}_{\rm ion}/M_h$, to be more uncertain. We plan to explore the physics relevant to these extreme epochs in future works. In addition, the models will be tested more stringently as high-redshift observational constraints from *JWST* and other instruments become available.

Furthermore, the EPS-based merger trees adopted in this dissertation work have been compared trees extracted from numerical simulations and the results shown to be in good agreement. However, the EPS algorithm has never been tested over the full halo mass and redshift ranges that are explored in this work, as there is currently no publicly available relevant suite of dark matter only simulations. We plan on running and analysing this suite of N-body simulations, and developing and validating a new and improved fast merger tree algorithm, in Yung et al. (in preparation).

Our analytic reionization model does not account for density fluctuations and clustering of sources across the Universe, which numerical simulations have shown can lead to an inhomogeneous and 'patchy' progression of reionization. Furthermore, our models do not self-consistently account for photoionization feedback (or 'squelching'). Other works have shown (Gnedin 2014; Mutch et al. 2016) that photoionization affects only galaxies in extremely low-mass halos, which we found have

a negligible contribution to reionization. Furthermore, post-reionization IGM temperature fluctuation can also be used to constrain the reionization history and this has been explored analytically (Furlanetto & Oh 2009) and with fully coupled radiation-hydrodynamic simulations (Wu et al. 2019).

Some sources that could be potential contributors to the total ionizing photon budget are not accounted for in this work. These include Population III stars, X-ray binaries, and AGN. Previous work has shown that Pop III stars are unlikely to dominate the reionizing photon budget (e.g. Ricotti & Ostriker 2004; Greif & Bromm 2006; Ahn et al. 2012; Paardekooper et al. 2013; Robertson et al. 2015), but they could make reionization more patchy, and are presumably important for polluting early halos with metals, which can then provide the seeds for dust and molecular hydrogen formation. This process is a critical component in our models, which is currently treated in a simplified way by adopting a metallicity 'floor' in all pristine halos. We further assume that significant cooling cannot occur in halos below the atomic cooling limit (10^4 K). While some cooling may occur at lower temperatures due to molecular hydrogen cooling or metal cooling, these are thought to be sub-dominant (Yoshida et al. 2004; Maio et al. 2010; Johnson et al. 2013; Wise et al. 2014; Xu et al. 2016; Jaacks et al. 2018). The contribution of early accreting black holes to reionization remains very uncertain, and we plan to investigate this in upcoming work. In addition to directly contributing ionizing photons through their hard spectrum, semi-analytic calculations have shown that X-rays produced by AGN can boost $f_{\rm esc}$ (Benson et al. 2013; Seiler et al. 2018). However, hydrodynamic simulations have shown that this effect is not significant (Trebitsch et al. 2018). We also note that the contribution from AGN to the total ionizing photon budget can become fairly significant near the completion of HI reionization (e.g. Dayal et al. 2020), and our results matching the BB13 emissivity constraints may imply an over-prediction of the contribution from galaxies.

Another caveat related to the observational constraints is that the estimates of $\tau_{\rm CMB}$ are highly covariant with other cosmological parameters, and are derived assuming a simple instantaneous reionization model. As additional constraints on $Q_{\rm HI}$ and the ionizing photon emissivity are obtained, and we gain a better understanding of the uncertainties on these measurements, these could be incorporated as additional constraints in a fitting procedure.

5.4.3 Constraining galaxy formation during the EoR with *JWST* and beyond

With both deep- and lensed-field NIRCam surveys anticipated to reach unprecedented detection limits, the extremely sensitive *JWST* is expected to directly detect and constrain the number density of faint galaxies up to $z \sim 10$. Furthermore, MIRI and NIRSpec will provide high-resolution spectroscopic follow-ups for the spectral features of these galaxies, which will put more robust constraints on $\xi_{\rm ion}$ and $f_{\rm esc}$. These measurements will allow us to further test and refine galaxy formation models

and to understand the physics that shapes galaxy properties at ultra-high redshift.

The coming decades promise great opportunities for further exploring the high-redshift Universe. The line-up of flagship instruments, include space-based *Euclid* (Racca et al. 2016) and *Wide-Field Infrared Survey Telescope* (*WFIRST*, Spergel et al. 2015), as well as the ground-based Large Synoptic Survey Telescope (LSST, LSST Science Collaboration 2017). These facilities are capable of surveying large areas, which is complementary to the small field-of-view of *JWST*. Furthermore, next generation facilities European Extremely Large Telescope (ELT, Gilmozzi & Spyromilio 2007), Thirty Meter Telescope (TMT, Sanders 2013), and Giant Magellan Telescope (GMT, Johns 2008) have the capability of doing spectroscopic follow-up on the expected large number of photometric detections. The flexibility of our model allows it to be easily adapted to made predictions for these instruments, and facilitate physical interpretation for future multi-instrument surveys. In addition, the Atacama Large Millimeter Array (ALMA) has the capability of detecting dust continuum as well as fine structure lines such as [CII] and [OII] of $z > 6$ galaxies. With the extended modelling framework presented in (Popping et al. 2019) coupled with our SAMs, we will also be able to make predictions for joint *JWST*–ALMA multi-tracer surveys.

Intensity mapping is a complementary approach that surveys large areas of the sky at relatively coarse angular resolution, potentially providing direct constraints on the conditions of the intergalactic hydrogen and indirect, collective constraints on high-redshift galaxy populations (Visbal & Loeb 2010; Visbal et al. 2011; Kovetz et al. 2017). Numerous intensity mapping experiments for HI, CO, CII, and Lyα are planned or underway, including BINGO (Battye et al. 2013), CHIME (Bandura et al. 2014), EXCLAIM (Padmanabhan 2019), HERA (DeBoer et al. 2017), HIRAX (Newburgh et al. 2016), Tianlai (Chen 2012), LOFAR (Patil et al. 2017), MeerKat (Pourtsidou 2018; Santos et al. 2017), CONCERTO (Serra, Doré, & Lagache 2016), PAPER (Parsons et al. 2010), etc., which together pave the way to future large-scale multi-tracer intensity mapping surveys. These observations can also be cross-correlated with galaxy surveys for a comprehensive view of the interaction between galaxies and the cosmic environment. The modelling framework presented here can also provide a powerful tool for efficiently producing physically self-consistent, multi-tracer predictions for intensity mapping experiments (Yang et al. in preparation).

Finally, improving radiative hydrodynamic simulations of early galaxy evolution (e.g. Finlator et al. 2018; Wu et al. 2019) will complement our approach by providing more physically motivated priors for our key physical parameters, and suggesting new parametrizations that connect quantities such as escape fraction to galaxy properties rather than redshift, as we have assumed here. Our approach provides a framework to bridge these detailed self-consistent models with upcoming deep and wide surveys to optimally constrain the physics of early galaxy formation.

5.5 Summary and Conclusions

In this work, we constructed a physically motivated, source-driven semi-analytic modelling pipeline that links galaxy formation to the subsequent reionization history using an analytic model for reionization. The galaxy formation model has been tested extensively and shown to match extremely well with observational constraints up to $z \sim 10$ in previous works, and we extended these predictions up to $z \sim 15$. We have calculated $\dot{N}_{\rm ion}$ self-consistently, accounting for the stellar age and metallicity distribution of the stellar population in each galaxies using state-of-the-art SPS models. We presented predictions for the ionizing emissivity, IGM neutral fraction, and Thomson optical depth to CMB throughout the Epoch of Reionization, and compared these to a wide range of observational constraints. In a controlled experiment, we isolated and quantified the effect of each of the major moving parts in the total ionizing photon budget. We also explored two different scenarios with a constant and a redshift-dependent $f_{\rm esc}$, and determined the required conditions for the predicted galaxy populations to reionize the Universe in the time frame require by IGM and CMB constraints. We explored the covariance of different model components (including $f_{\rm esc}$ and the efficiency of stellar feedback) using MCMC.

We summarize our main conclusions below.

1. Using a well-tested physical galaxy formation model, which was calibrated only to $z \sim 0$ observations and has been shown to well-reproduce observed distributions from $z \sim 4$–10, we provide predictions for rest-frame UV luminosity functions and ionizing photon production rate for galaxies up to $z = 15$.

2. Adopting a non-evolving escape fraction of $\sim 20\%$, the galaxy population predicted by our model yields sufficient amounts of ionizing radiation to be consistent with constraints from the Thomson optical depth $\tau_{\rm CMB}$. However, this model is in tension with low-redshift Lyα observations on the IGM neutral fraction and observational constraints on the ionizing emissivity at $2 \lesssim z \lesssim 6$.

3. We performed a number of controlled experiments to explore the impacts on the reionization history of varying the three main model components that influence the total ionizing photon budget, including the abundance of low-mass galaxies, intrinsic ionizing photon production rate, and LyC escape fraction. We find that the uncertainty on estimates of the total LyC emissivity is dominated by uncertainties on $f_{\rm esc}$, with the strength of stellar feedback being the second most important factor.

4. We used MCMC to explore the covariance in these two parameters ($f_{\rm esc}$ and $\alpha_{\rm rh}$, which

parametrizes the efficiency of stellar feedback in low-mass halos). We parametrized the population averaged $f_{\rm esc}$ as a function of redshift, and jointly constrained these parameters along with $\alpha_{\rm rh}$ using constraints from Lyα forest observations and $\tau_{\rm CMB}$ measurements. We found that a 'population-averaged' escape fraction that mildly increases from $\sim 4\%$ to $\sim 29\%$ between $z \sim 4$ to 15 satisfies both constraints.

5. We presented predictions for the fraction of ionizing photons produced by galaxies of different rest-UV luminosity as a function of redshift, and for the fraction of the total ionizing photon budget sourced by galaxy populations that will be observable in upcoming surveys with *JWST*. At $z \sim 7$–8, where we predict the Universe to be about 50% reionized by volume, we predict that *JWST* will be able to detect the sources of 60–70% of the reionizing photons in a deep survey, and up to $\sim 90\%$ in an ultra-deep lensed survey.

Chapter 6
Conclusion

In this dissertation work, we constructed a semi-analytic modelling pipeline to predict and investigate the properties of high-redshift galaxies and their impact on the subsequent cosmic reionization history. This pipeline is the first of its kind to have taken a fully semi-analytic approach to constructing dark matter halo merger trees, modelling galaxy formation, and tracking the volume-averaged progression of the reionization of intergalactic hydrogen. This set-up enables self-consistent explorations of the connections between galaxy formation physics and a wide variety of cosmological observables. Taking advantage of the pipeline's efficiency, we also conducted controlled experiments by systematically varying a set of key physical parameters. These results are critical to quantifying the impact of the sensitivities of our predictions to uncertainties in current physical models and assessing the capabilities of future instruments on constraining galaxy formation physics. We also demonstrated in this dissertation that semi-analytic models can be an extremely useful companion to large survey programs and have applied some of our predictions to provide insights to some of the main open questions that will be address by future flagship telescopes.

In the first half of this thesis, we extensively examined the galaxy populations predicted for very high redshifts ($z \sim 4$–10), including the statistical distributions and scaling relations for their physical and photometric properties. While exploring the effects from a variety of star formation-related models, we pinpointed the set of model components, configurations, and physical parameters that are required to reproduce the observed distribution of high-redshift galaxies, which includes a H_2-based star formation relation that becomes more efficient in H_2-dense gas and a specific combination of SN feedback strength and molecular gas depletion time. We find that the galaxy formation physics we learned from observations in the nearby universe are quite sufficient to explain the observed evolution for galaxies during the EoR. Our simulated galaxy populations also span an extremely wide mass range, wider than most numerical and semi-analytic simulations currently available, which allows us to make predictions for low-mass objects that fall below the current detection limits. We explored a range of possible outcomes amid the current modelling uncertainties and concluded that the faint-end slope of UV LFs is mainly sensitive to the scaling of the mass-loading factor for stellar driven winds with halo or galaxy properties. We also found that star formation physics and gas depletion

time have dominant effects on the formation of massive galaxies, but the observed abundance of bright galaxies is subject to degenerate effects from dust attenuation. These results can be used to guide physical interpretation of future observations.

In the second half of this thesis, we implemented and tested new model components to expand the model's capabilities for making a variety of EoR-related predictions. With the crucial component that calculates the production rate of ionizing photons self-consistently based on the star formation and chemical enrichment histories of the predicted galaxy populations, we are able to estimate the total ionizing photon budget supplied to reionize the IGM by galaxies, while accounting for variance among populations of galaxies and distinguishing the contributions from galaxies of different luminosities or masses. The results from our physically motivated model painted a coherent picture that high-redshift galaxies predicted with a set of well-tested physics are able to produce sufficient ionizing photons to reionize the Universe in the time frame required by the latest IGM and CMB observations. We also find that low-mass galaxies are more efficient at producing ionizing photons than massive galaxies and high-redshift galaxies are more efficient than their low-redshift counterparts.

We applied some of our results to create *forecasts* specifically for the highly-anticipated *JWST*, which possesses unprecedented IR sensitivity and is expected to directly detect galaxies during the EoR. Based on the publicly available instrument and broadband filter specifications, we showed that *JWST* will be able to put constraints on the number density of faint galaxies down to a rest-frame $M_{\rm UV} \sim -15.68$ or $\log M_* \sim 6.44$ at $z \sim 10$, assuming some typical deep-field survey configurations. These future observations will provide constraints that help understand the build-up of galaxies in the early universe, as well as the strength and evolution of stellar feedback effect and other relevant physical processes. We also predicted that *JWST* in deep-field surveys will be able to detect sources that are accountable for $\sim 60\%$ to $\sim 70\%$ of the total ionizing photon budget, and up to $\sim 90\%$ in an ultra-deep lensed survey. The pipeline and peripheral tools used to create these predictions can be easily adapted to serve other instruments, and can be used to support the scientific rationales for the design of instruments or survey programs, optimize observing strategies, and facilitate physical interpretation of results. Modelling results and predictions presented in this dissertation have been made accessible to the astronomical community and are being used by a number of collaborations on exploring *JWST* and *WFIRST* science.

6.1 Future Work and Long-Term Development

The work presented in this dissertation will serve as the basis for future development of *Semi-Analytic Models with Ultrahigh-Redshift Astrophysical Interplay* (***SAMURAI***), which is envisioned to be

a modelling enterprise that will bring continuous, long-term advancements to the well-established Santa Cruz SAM with special focus on physical processes that are known to have significant influences at very high redshifts ($z \gtrsim 10$). This will equip our model with tools to tackle other astrophysical conundrums and to keep up with the rapidly evolving field. This proposed project has been accepted by the NASA Postdoctoral Program (NPP) beginning in October, 2020.

Despite the current success of modelling the evolution of observed high-redshift galaxy populations and providing predictions for low-mass galaxies to be observed by *JWST* and beyond, we must acknowledge that many physical processes that have strong influences on these primordial galaxies are not currently incorporated in the model. As shown in numerical and (semi-)analytic simulations, the co-evolution of the first stars and the first black holes (BHs) have some leading, far-reaching effects that propagate through time and have profound influences on the evolution of other galaxy components. In the current paradigm of structure formation, the first stars formed in pristine halos and were the primary source of metals, which are then distributed across their host halos via SN explosions. This very first batch of metals enables hot gas to cool efficiently for the first time and kick off the cycles of metal production and star formation. The remnants of the first stars then become one of the main channels seeding the first BHs, along with others that are seeded by massive gas clouds that directly collapse under self-gravitation. These seeds grow over time into supermassive BHs and quasars via accretions and mergers and co-evolve with the morphology and physical features of galaxies (e.g. Somerville et al. 2008).

By modelling the first stars using an approach similar to the one developed by Visbal et al. (2018), this proposed model will be able to predict where and in which halos these direct collapse BHs can form, provided that some of the conditions required are already provided by the existing modelling framework. One of the powerful aspects of the project would be the ability to simultaneously include BH seeds from first stars, which are relatively low mass but more prevalent, and seeds from direct collapse, which are more massive but probably rare. It is evident that both of these seeding channels are required as inferred by current observations. However, incorporating both seeding channels in cosmological-scale hydrodynamic simulations is highly non-trivial even with the most powerful machines currently available, as it requires extremely high mass-resolution to properly resolve the halos that host the seeds of luminous quasars and extremely large simulated volumes to sample these very rare objects. This proposed model will be the first to fully explore both seeding mechanisms in a cosmological context and provide predictions for new physical quantities to be detected in the future. This model advancement will also yield more robust predictions for the properties of primordial galaxies that can be observed with *JWST* and *WFIRST* (Volonteri et al. 2017).

This timely addition to the modelling pipeline will allow it to address the physical origin of the large fraction of optically obscured AGNs like the ones discovered recently (e.g. LaMassa et al. 2019), and how would this potentially large population of AGN fit into the current picture of cosmic reionization explored in Chapter 5. There are also puzzles remaining in the physical origin of the billion-solar-mass BHs needed to power those massive AGNs found by $z \sim 7$, and perhaps the seeding of BHs may be the key to understanding the development of these AGNs. Other relevant applications would be predicting properties of 'fossil' stars and BH occupation fractions in nearby dwarf galaxies. Furthermore, this will also enable predictions for black hole accretions and mergers, which will be detectable in x-ray and gravitational waves by a number of approved flagship instruments, such as the *Advanced Telescope for High ENergy Astrophysics* (*ATHENA*) and the Laser Interferometer Space Antenna (LISA). These will produce predictions for new physical quantities that can be cross-correlated with the wide range of existing predictions and form the basis of a unified platform for the physical interpretations for some highly anticipated multi-messenger surveys.

Appendix A
Data Release and Repository

The data products and mock galaxy catalogues associated with the work presented in this thesis are released and made accessible to the astronomical community through a permanent location at https://www.simonsfoundation.org/semi-analytic-forecasts-for-jwst/, which is hosted and administered by the Center for Computational Astrophysics at the Flatiron Institute in New York.

Tabulated distribution functions include:

- Rest-frame UV luminosity functions for various SF models, with and without dust attenuation
- Observed-frame apparent magnitude functions for full set of NIRCam broadband filters
- Stellar mass functions
- Star formation rate functions
- Cold gas and molecular mass functions
- Scaling relations among selected star formation-related observable and physical properties
- cosmic ionizing photon productivity for model variants
- ionizing photon production rate and efficiency as a function of halo mass
- IGM reionization history for model variants

Full object catalogues with a comprehensive list of physical and observable properties predicted at $z = 4$–15 with our fiducial model are released via the online portal.

Appendix B
Re-calibration for Planck Cosmology

In this appendix, we briefly summarize the calibration of our model after updating the cosmological parameters to values consistent with recent constraints from Planck. The free parameters in our model are calibrated by hand such that the outputs from our fiducial model match observations at $z \sim 0$. The calibration quantities are the stellar-to-halo mass ratio and stellar mass function, stellar mass-metallicity relation, cold gas fraction versus stellar mass relation for disk-dominated galaxies, and the black hole mass vs. bulge mass relation. We adopt observational constraints for these quantities from Rodríguez-Puebla et al. (2017, and references therein), Bernardi et al. (2013), Gallazzi et al. (2005), Peeples et al. (2014), Calette et al. (2018), and McConnell & Ma (2013). We show our model outputs compared with these observational constraints in Fig. B.1. In addition, we show two quantities used as a cross-check but not used directly in the calibration: cold gas phase mass-metallicity relation (lower-left panel in Fig. B.1) and the H_2 mass function (Fig. B.2), along with observational constraints from Obreschkow & Rawlings (2009b), Keres et al. (2003), Andrews & Martini (2013), Zahid et al. (2013), and Boselli et al. (2014). For scaling relations, the blue solid line marks the median and the dashed lines mark the $16^{\rm th}$ and $84^{\rm th}$ percentile for central galaxies. Both central and satellite galaxies are included in the distribution functions (such as the SMF and H_2 mass function). SPT15 has shown that the different SF recipes produce results that converge at $z \sim 0$. The re-calibration for updated cosmology have uniform effects across the models and therefore should not qualitatively change the results among models. Hence, we refer reader to such work for the relative differences among SF models at low redshifts.

Traditionally, there has been tension in the predicted physical properties in galaxy formation models. In one case, matching the stellar fraction closely tends to lead to an excess in gas fraction. Since the underlying star formation efficiency that has immediate effects on stellar fraction is directly related to the production of metals, which our multiphase-gas partitioning, star formation, and dust attenuation recipes are extremely sensitive to, calibrating our model to simultaneously match the observed gas fraction, stellar metallicity, and stellar fraction is very challenging. The uncertainty of about a factor of two in the observed normalization of the Kennicutt SF law and the observed gas fractions provides us with some leeway for calibrating our model. After carrying out multiple tests

with different configurations, we find that leaving the SF timescale $\tau_{*,0}$ close to unity yields the best results, with gas fractions slightly higher than values reported by Calette et al. (2018). If we strictly enforce the gas fraction to match observations by decreasing $\tau_{*,0}$, we will also need to compensate for that by increasing AGN feedback to keep the model from overproducing massive galaxies.

While we configure the SF timescale and AGN feedback to fit the massive end of the stellar-to-halo mass ratio and stellar mass function, the faint populations seem to be more sensitive to the SN feedback slope $\alpha_{\rm rh}$. Structure forms earlier in the Planck cosmology relative to the WMAP-5 cosmology used in our previous work, leading to a higher number density of low-mass dark matter halos at early times. This requires us to increase $\alpha_{\rm rh}$ to suppress the formation of low-mass galaxies in order to match both the stellar-to-halo mass ratio and stellar mass function, which comes with a side effect of further steepening the mass-metallicity relation. As discussed in previous works SPT15, this tension, along with other discrepancies at intermediate redshifts (White et al. 2015) hint that our rather simple recipe for stellar feedback needs to be revised. Moreover, since the abundance of low-mass galaxies is very sensitive to the choice of $\alpha_{\rm rh}$, deviating from the calibrated value would certainly would certainly raise tension with observational constraints at $z \sim 0$. The impact of alternative $\alpha_{\rm rh}$ values at $z \sim 0$ has been examined in (White et al. 2015), which they found that making alternate assumptions for ejected mass would lead to a ~ 0.25 dex changes relative to the fiducial model. Some of the values, including the ones disfavored by local observations are explored in this work in effort to quantify its effect on the low-mass galaxy populations. The differences among the three star formation recipes presented in this work have been explored in SPT15.

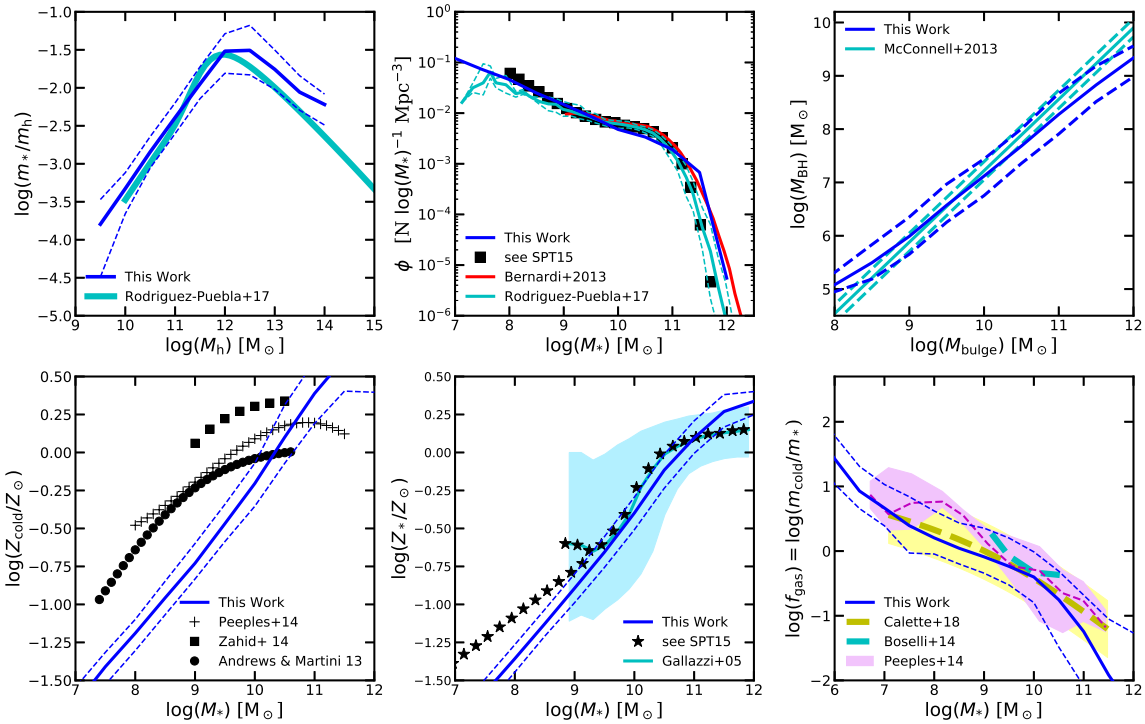

Figure B.1 Outputs at $z = 0$ from our fiducial model compared to the observational constraints on *top row from left to right*: stellar-to-halo mass ratio, stellar mass function, and $M_{\rm BH}$-$M_{\rm bulge}$ relation; *bottom row from left to right*: cold gas metallicity, stellar metallicity, and gas fraction reported by various studies. See text for full details.

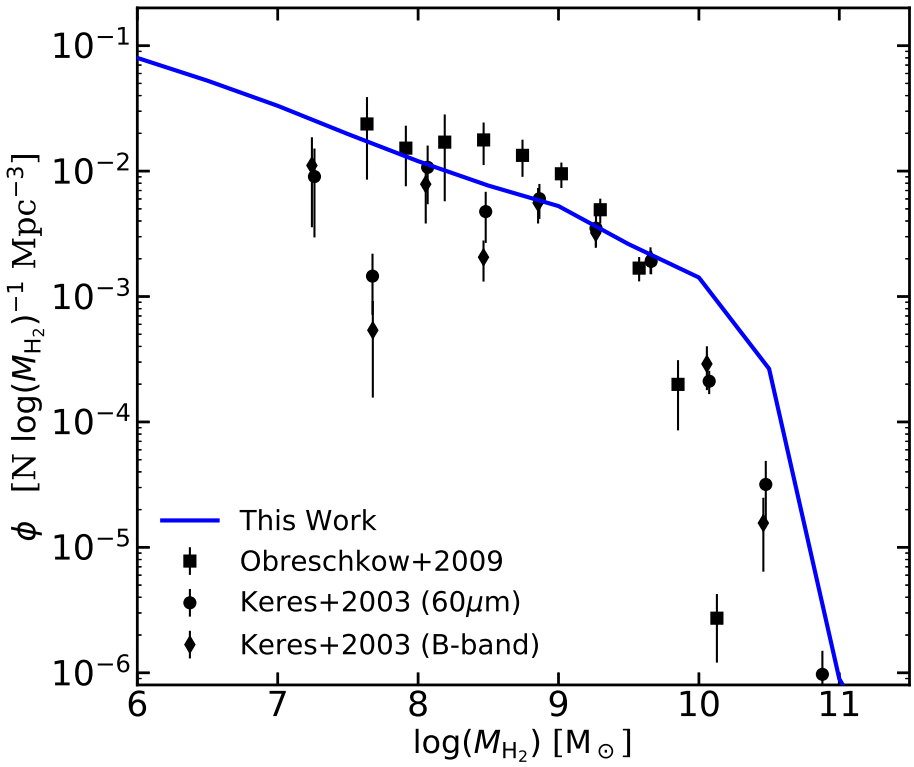

Figure B.2 H_2 mass function at $z = 0$ from our fiducial model compared to the observational constraints presented in Keres et al. (2003) and Obreschkow & Rawlings (2009b).

Appendix C

Testing the Halo Mass Function

In this work, we use the fitting functions for halo mass functions provided by Rodríguez-Puebla et al. (2016) that is fitted to the Bolshoi-Planck simulations (Klypin et al. 2016). However, the mass resolution from Bolshoi-Planck simulation is well above $V_{\rm vir} \sim 20\ \mathrm{km\,s^{-1}}$which our grid of root halos reaches. To check the validity of these fitting functions for HMF, we used are compared to very high resolution, small box simulations that are similar to the ones presented in Visbal et al. (2018). A comparison between the halo mass functions and the numerical simulations is shown in fig. C.1.

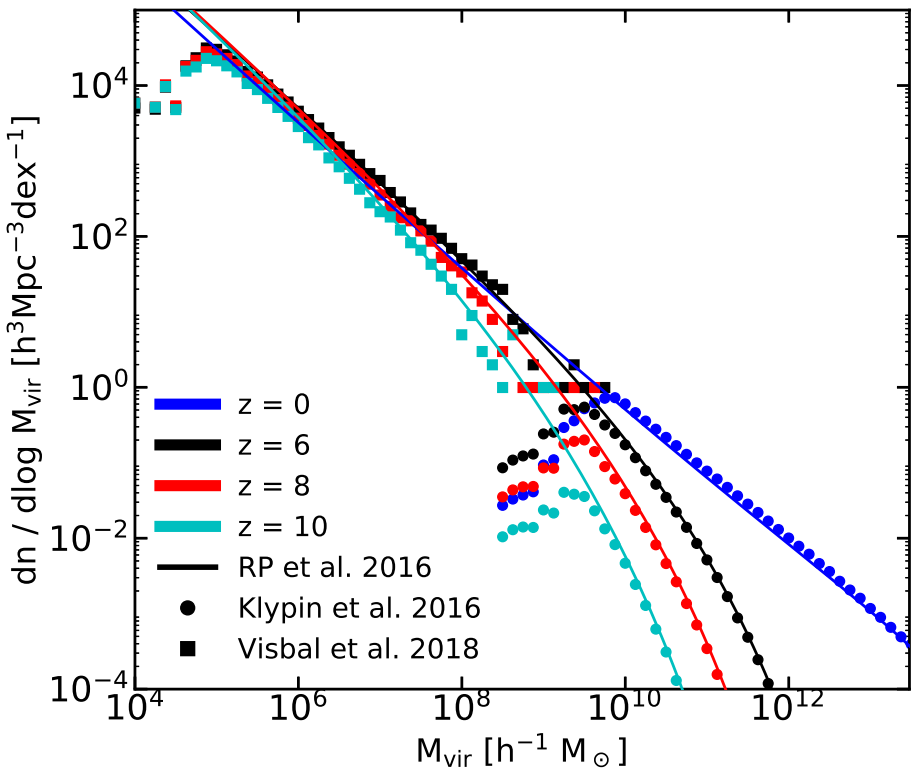

Figure C.1 A comparison of the HMF fitting functions from Rodríguez-Puebla et al. (2016, solid line) to numerical simulations from Klypin et al. (2016, circle marker) and Visbal et al. (2018, square marker).Plot elements are color-matched by redshift.

Appendix D

Fitting Functions for Selected Scaling Relations

The fitting parameters for the medians of selected scaling relations presented in 3.4 are provided in Table D.1. As noted in the text, the correlations break down in the bright, massive galaxies due to the effect of dust attenuation, which is not accounted for in the fitting. Bright galaxies with $m_{\mathrm{m200W}} < 27$ and $M_{\mathrm{UV}} < -24$ are excluded from our fits. Galaxies with $\log(M_*/\mathrm{M}_\odot) < 6$ are also excluded as the scatter increases below this mass (see Fig. 3.16, 3.17 3.18, 3.19, and 3.20. In general, the residuals on these fits range from a few percent to $\sim 15\%$.

Table D.1 Fitting parameters for the medians of selected scaling relations. For each pair of X-Y relation, the fitting parameters for $Y = Ax + B$ are given below.

	m_{F200W}-M_{UV}		m_{F200W}-M_*		m_{F200W}-SFR		M_{UV}-M_*		M_*-SFR	
z	A	B	A	B	A	B	A	B	A	B
4	0.9048	-42.9055	-0.4033	19.4533	-0.3746	9.9795	-0.4420	0.2893	0.9692	-8.3461
5	0.9161	-43.9573	-0.4000	19.4269	-0.3829	10.5306	-0.4233	0.3931	0.9828	-8.2576
6	0.9139	-44.1058	-0.3991	19.3298	-0.3828	10.6324	-0.4244	0.2154	0.9679	-7.9788
7	0.9088	-44.0984	-0.3988	19.2443	-0.3823	10.6624	-0.4266	0.0423	0.9472	-7.7057
8	0.9200	-44.5426	-0.4058	19.3903	-0.3950	11.0885	-0.4279	-0.1031	0.9532	-7.6551
9	0.9274	-44.8604	-0.3983	19.0822	-0.3919	10.9837	-0.4242	-0.1430	0.9543	-7.6178
10	0.9120	-44.4035	-0.3880	18.6637	-0.3856	10.7586	-0.4208	-0.1673	0.9513	-7.5384

Appendix E
Impact of Adopting Updated Cosmological Parameters

During the course of recalibrating the model parameters, we found that changing from the WMAP cosmology to the Planck cosmology alone has some significant effects on our predictions for high-redshift galaxy populations. For each iteration of our model (e.g. S08; SPT15; Chapter **??** of this thesis), the parameters are tuned to match a subset of $z \sim 0$ observables. Although the specific observations used for calibration have changed throughout the years as the available observational constraints improved, the qualitative evolution of galaxy populations remained similar. However, as we move to $z \gtrsim 4$, the difference arising from the adopted cosmology becomes more noticable. Furthermore, different works have used various approximate fitting functions to estimate the abundance of dark matter halos as a function of mass and redshift. For example, S08 and SPT15 used the Sheth, Mo, & Tormen (2001) fitting function, while in this work we use the updated fitting function from Rodríguez-Puebla et al. (2016). In Fig. E.1 we show the effects of both fitting function choice and cosmology on the halo mass function at $z \sim 10$–0. These factors can lead to differences in halo abundance of a factor of two or more.

In this appendix, we focus on results at $z = 4$ from this work and compare them to results from S08 and SPT15. The S08 models used the WMAP3 cosmology, and results from these models are shown in (Porter et al. 2014), the Lu et al. (2014a) SAM comparison and the Song et al. (2016) and Duncan et al. (2014) comparison to observations. The updated multiphase models presented in Somerville et al. (2015) incorporate the same gas-partitioning and H_2-based SF recipes used here, but adopted the WMAP7 cosmology (see Table E.1).

We compare the SMFs from S08 and the non-multiphase version of SPT15, and then the SMFs from the multiphase version of SPT15 and from this work. Our goal is to show the weak residual effects of the choice of cosmology on the redshift evolution of galaxy properties even *after* model recalibration.

The Planck cosmology predicts higher abundances of halos of all masses and at all redshifts than WMAP, although not by the same factor at all redshifts and masses, as can be seen in Fig. E.1. In Chapter 2 of this thesis, we describe how we recalibrated the parameters controlling the physical processes in the SAM to produce nearly identical results for our calibration quantities at $z \sim 0$

(Appendix C in Chapter 3). As shown in Fig. E.2, the updated model produces a very small excess of bright galaxies compared to the previously published (SPT15) model. The residual difference at high redshift due to cosmology is also weakly dependent on the way that physical processes are modeled in the SAM, as can be seen from the models with different star formation recipes run within each of the two cosmologies.

Table E.1 Summary of cosmological parameters used in previous models.

	WMAP3 (S08)	WMAP5 (SPT15)	Planck 2015 (This Work)
Ω_m	0.2383	0.28	0.308
Ω_Λ	0.7617	0.72	0.692
H_0	73.2	70.0	67.8
f_b	0.1746	0.1658	0.1578
σ_8	0.761	0.81	0.831
n_s	0.958	0.96	0.9665

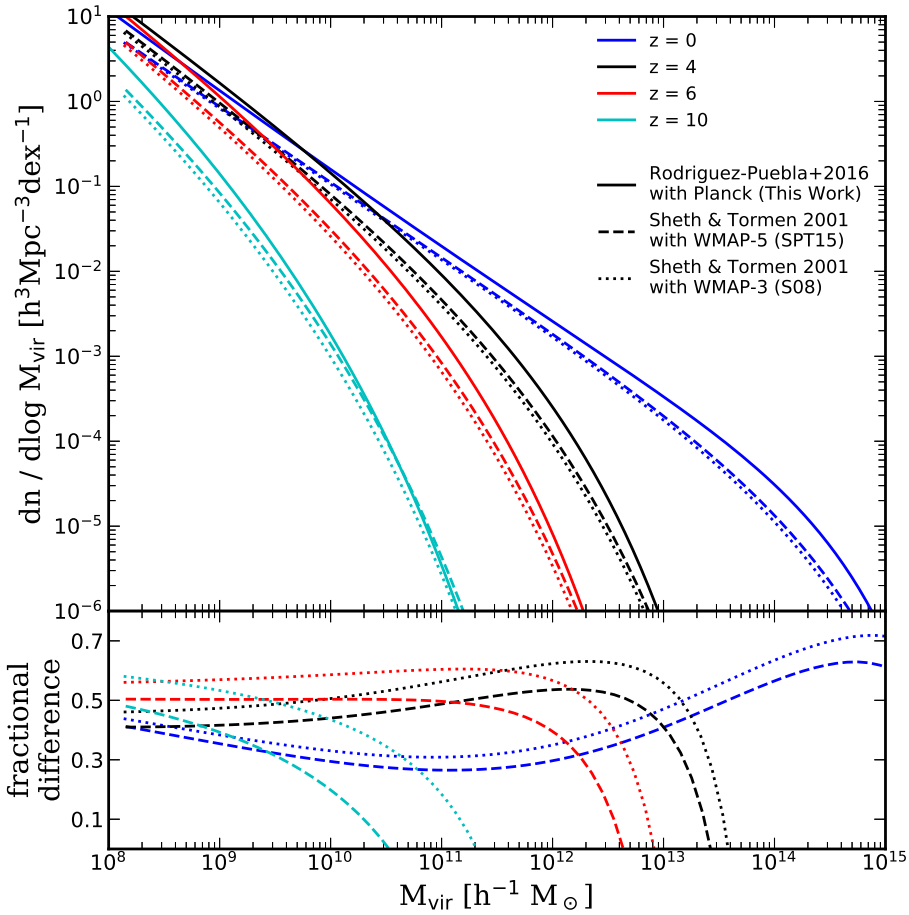

Figure E.1 Halo mass functions from the fitting formulae of Sheth et al. (2001) and Rodríguez-Puebla et al. (2016) for different choices of cosmology: Planck (This Work, solid line), WMAP-7 (SPT15, dashed line), WMAP-3(S08, dotted line). The bottom panel shows the fractional difference relative to the RP16 Planck HMFs. The Planck cosmology generally predicts higher abundances of halos than WMAP cosmologies over a broad range of mass and redshift.

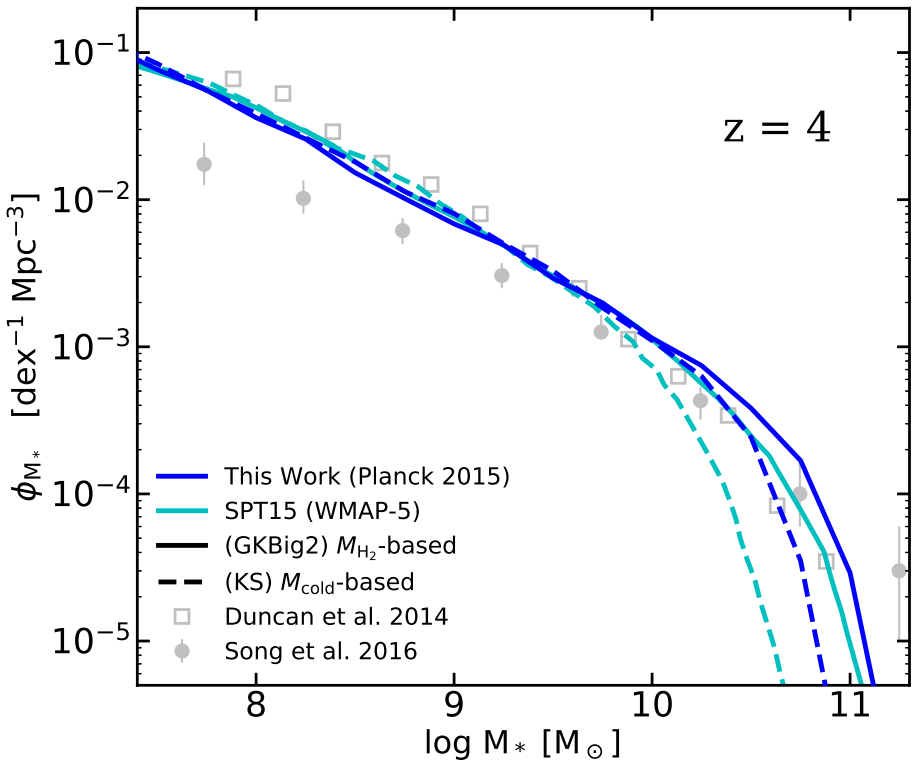

Figure E.2 Stellar mass functions at $z = 4$ predicted by our SAM with different cosmologies and star formation recipes, *after* retuning to match observational constraints at $z \sim 0$. These data are color-matched for the cosmology, where blue and cyan represent Planck 2015 and WMAP-5; and are style-matched for the star formation model, where solid and dashed lines represent the multiphase H_2-based GKBig2 and the non-multiphase $M_{\rm cold}$-based KS model. This figure demonstrates that our predictions at high redshift show a weak dependence on cosmology after recalibration.

Appendix F

HMF Fitting Parameters for the Extended Redshift Range

We adopted the HMF parametrization from Tinker et al. (2008) with parameters calibrated to the Bolshoi-Planck simulation from the MultiDark suite (Rodríguez-Puebla et al. 2016; Klypin et al. 2016). The comoving number density of halos of mass between $M_{\rm vir} + dM_{\rm vir}$ is given by

$$\frac{dn_{\rm h}}{dM_{\rm vir}} = f(\sigma)\frac{\rho_m}{M_{\rm vir}^2}\left|\frac{d\ln\sigma^{-1}}{d\ln M_{\rm vir}}\right|, \tag{F.1}$$

where ρ_m is the critical matter density in the Universe, σ is the amplitude of the perturbations, and $f(\sigma)$ is called the halo multiplicity function, which takes the form of

$$f(\sigma) = A\left[\left(\frac{\sigma}{b}\right)^{-a}\right]e^{-c/a^2}, \tag{F.2}$$

where A, a, b, and c are free parameters. In this work, as shown in Fig. G.1, we recalibrate these parameters to match the HMF constraints between $z = 11 - 15$ from the Bolshoi-Planck simulation and from Visbal et al. (2018). These parameters are presented in Table F.1.

Table F.1 Fitting parameters for $f(\sigma)$ parameters that produces the HMF at $z = 11 - 15$ used throughout this work as shown in fig. G.1.

z	A	a	b	c
11	0.1668	0.9823	1.100	1.0938
12	0.1468	0.9823	1.000	1.0938
13	0.1468	0.9823	0.900	1.0938
14	0.1268	0.9823	0.750	1.0938
15	0.1268	0.5523	0.600	1.1238

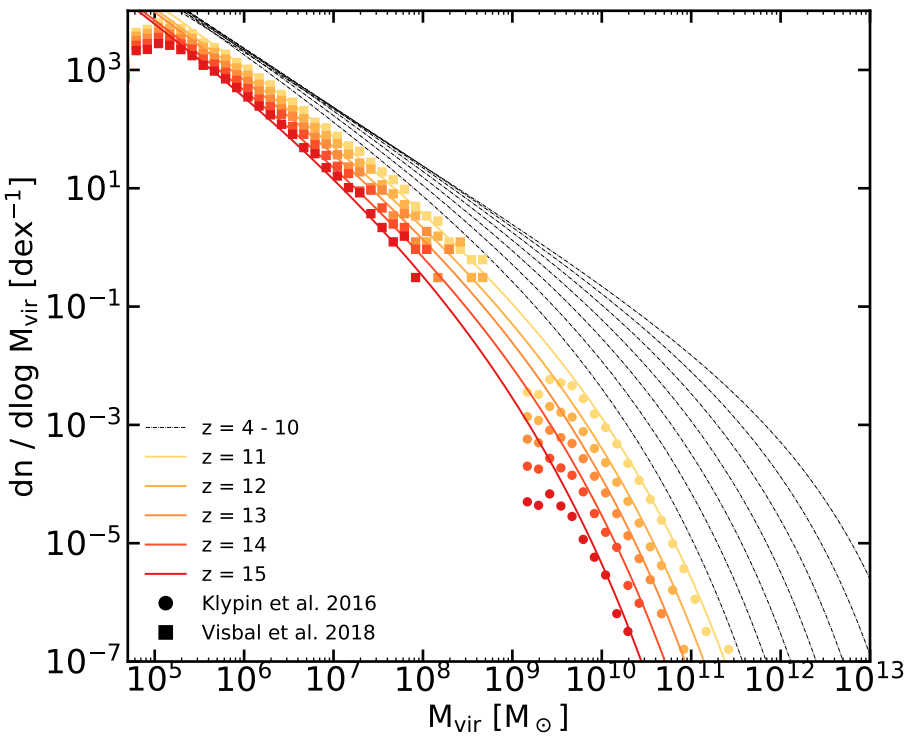

Figure F.1 The coloured lines, from light to dark, show the HMF fitting functions adopted for the extended redshift range $z = 11$–15. n-body simulation predictions from Klypin et al. (2016) and Visbal et al. (2018) are shown in matching colour for each redshift for comparison. The light grey dot-dashed lines show HMF fitting functions used for $z = 4$–10 to guide the eye (see Chapter 2 for detail).

Appendix G

Logo Associated with the Project

This project logo we designed made its presence in numerous talks, posters, and webpages, which played a significant role in representing this work and in promoting the content of this work to the scientific community.

Figure G.1 The *Semi-analytic forecasts for JWST* project logo.

Bibliography

Ahn, K., Iliev, I. T., Shapiro, P. R., Mellema, G., Koda, J., & Mao, Y. 2012, ApJ, 756, L16

Alpher, R. A., & Herman, R. 1948, Nature, 162, 774

Anderson, L., Governato, F., Karcher, M., Quinn, T., & Wadsley, J. 2017, MNRAS, 468, 4077

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Arimoto, N., Sofue, Y., & Tsujimoto, T. 1996, PASJ, 48, 275

Atek, H., Richard, J., Jauzac, M., et al. 2015, ApJ, 814, 69

Atek, H., Richard, J., Kneib, J.-P., & Schaerer, D. 2018, MNRAS, 479, 5184

Aubourg, É., Bailey, S., Bautista, J. E., et al. 2015, Phys. Rev. D, 92, 123516

Bandura, K., Addison, G. E., Amiri, M., et al. 2014, Proc. SPIE, 9145, 914522

Barrow, K. S. S., Wise, J. H., Norman, M. L., O'Shea, B. W., & Xu, H. 2017, MNRAS, 469, 4863

Battye, R. A., Browne, I. W., Dickinson, C., Heron, G., Maffei, B., & Pourtsidou, A. 2013, MNRAS, 434, 1239

Becker, G. D., & Bolton, J. S. 2013, MNRAS, 436, 1023

Becker, R. H., Fan, X., White, R. L., et al. 2001, AJ, 122, 2850

Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729

Behroozi, P., Wechsler, R. H., Hearin, A. P., & Conroy, C. 2019, MNRAS, 488, 3143

Behroozi, P. S., & Silk, J. 2015, ApJ, 799, 32

Behroozi, P. S., Wechsler, R. H., & Conroy, C. 2013, ApJ, 770, 57

Benson, A., Venkatesan, A., & Shull, J. M. 2013, ApJ, 770, 76

Benson, a. J., Frenk, C. S., Lacey, C. G., Baugh, C. M., & Cole, S. 2002a, MNRAS, 333, 177

Benson, a. J., Lacey, C. G., Baugh, C. M., Cole, S., & Frenk, C. S. 2002b, MNRAS, 333, 156

Bernardi, M., Meert, A., Sheth, R. K., Vikram, V., Huertas-Company, M., Mei, S., & Shankar, F. 2013, MNRAS, 436, 697

Bertelli, G., Bressan, A., Chiosi, C., Fagotto, F., & Nasi, E. 1994, Astron. Astrophys. Suppl. Ser., 106, 275

Bertone, G., Hooper, D., & Silk, J. 2005, Phys. Rep., 405, 279

Bigiel, F., Leroy, A., Walter, F., Brinks, E., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2846

Bigiel, F., Leroy, A. K., Walter, F., et al. 2011, ApJ, 730, L13

Bisigello, L., Caputi, K. I., Colina, L., et al. 2016, ApJS, 227, 19

Bisigello, L., Caputi, K. I., Colina, L., Le Fèvre, O., Nørgaard-Nielsen, H. U., Pérez-González, P. G., van der Werf, P., Ilbert, O., Grogin, N., & Koekemoer, A. 2017, ApJS, 231, 3

Blumenthal, G. R., Faber, S. M., Primack, J. R., & Rees, M. J. 1984, Nature, 311, 517

Bolatto, A. D., Leroy, A. K., Jameson, K., et al. 2011, ApJ, 741, 12

Bolatto, A. D., Leroy, A. K., Rosolowsky, E., Walter, F., & Blitz, L. 2008, ApJ, 686, 948

Bolton, J. S., & Haehnelt, M. G. 2007, MNRAS, 382, 325

Bolton, J. S., Haehnelt, M. G., Warren, S. J., Hewett, P. C., Mortlock, D. J., Venemans, B. P., McMahon, R. G., & Simpson, C. 2011, MNRAS, 416, L70

Bond, J. R., Cole, S., Efstathiou, G., & Kaiser, N. 1991, ApJ, 379, 440

Boselli, A., Cortese, L., & Boquien, M. 2014, A&A, 564, A65

Bouwens, R. J., Aravena, M., Decarli, R., et al. 2016a, ApJ, 833, 72

Bouwens, R. J., Bradley, L., Zitrin, A., et al. 2014a, ApJ, 795, 126

Bouwens, R. J., Illingworth, G. D., Oesch, P. a., Caruana, J., Holwerda, B., Smit, R., & Wilkins, S. 2015a, ApJ, 811, 140

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2011, ApJ, 737, 90

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Labbé, I., van Dokkum, P. G., Trenti, M., Franx, M., Smit, R., Gonzalez, V., & Magee, D. 2014b, ApJ, 793, 115

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2010, ApJ, 709, L133

—. 2015b, ApJ, 803, 34

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Trenti, M., Labbé, I., Franx, M., Stiavelli, M., Carollo, C. M., van Dokkum, P., & Magee, D. 2012, ApJ, 752, L5

Bouwens, R. J., Oesch, P. A., Illingworth, G. D., Ellis, R. S., & Stefanon, M. 2017, ApJ, 843, 129

Bouwens, R. J., Oesch, P. A., Labbé, I., et al. 2016b, ApJ, 830, 67

Bouwens, R. J., Smit, R., Labbé, I., Franx, M., Caruana, J., Oesch, P., Stefanon, M., & Rasappu, N. 2016c, ApJ, 831, 176

Bower, R. G. 1991, MNRAS, 248, 332

Bowler, R. A. A., Dunlop, J. S., McLure, R. J., et al. 2015, MNRAS, 452, 1817

—. 2014, MNRAS, 440, 2810

Boylan-Kolchin, M., Bullock, J. S., & Garrison-Kimmel, S. 2014, MNRAS, 443, 44

Boylan-Kolchin, M., Weisz, D. R., Bullock, J. S., & Cooper, M. C. 2016, MNRAS, 462, L51

Boylan-Kolchin, M., Weisz, D. R., Johnson, B. D., Bullock, J. S., Conroy, C., & Fitts, A. 2015, MNRAS, 453, 1503

Brennan, R., Choi, E., Somerville, R. S., Hirschmann, M., Naab, T., & Ostriker, J. P. 2018, ApJ, 860, 14

Brinchmann, J., Charlot, S., White, S. D. M., Tremonti, C., Kauffmann, G., Heckman, T., & Brinkmann, J. 2004, MNRAS, 351, 1151

Brinchmann, J., & Ellis, R. S. 2000, ApJ, 536, L77

Bromm, V., Coppi, P. S., & Larson, R. B. 1999, ApJ, 527, L5

Bruzual, A. G. 1983, ApJ, 273, 105

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Bruzual, G. A., & Charlot, S. 1993, ApJ, 405, 538

Bryan, G. L., & Norman, M. L. 1997, ASP Conf. Ser. 123, Comput. Astrophys. ed. D. A. Clarke M. Fall (San Fr. ASP), 363

—. 1999, in Workshop on Structured Adaptive Mesh Refinement Grid Methods, ed. N. Chrisochoides (IMA Volumes in Mathematics 117), 165

Bullock, J. S., Kravtsov, A. V., & Weinberg, D. H. 2000, ApJ, 539, 517

Cabezón, R. M., García-Senz, D., & Figueira, J. 2017, A&A, 606, A78

Calette, A. R., Avila-Reese, V., Rodríguez-Puebla, A., Hernández-Toledo, H., & Papastergis, E. 2018, Rev. Mex. Astron. y Astrofísica, 54, 443

Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koornneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682

Carucci, I. P., & Corasaniti, P.-S. 2019, Phys. Rev. D, 99, 023518

Castellano, M., Fontana, A., Boutsia, K., et al. 2010, A&A, 511, A20

Castellano, M., Yue (), B., Ferrara, A., et al. 2016, ApJ, 823, L40

Ceverino, D., Glover, S. C. O., & Klessen, R. S. 2017, MNRAS, 470, 2791

Ceverino, D., Klessen, R. S., & Glover, S. C. O. 2018, MNRAS, 480, 4842

—. 2019, MNRAS, 484, 1366

Chabrier, G. 2003a, PASP, 115, 763

—. 2003b, ApJ, 586, L133

Charlot, S., & Bruzual, A. G. 1991, ApJ, 367, 126

Charlot, S., & Fall, S. M. 2000, ApJ, 539, 718

Chen, C.-C., Smail, I., Swinbank, A. M., et al. 2015, ApJ, 799, 194

Chen, X. 2012, Int. J. Mod. Phys. Conf. Ser., 12, 256

Chisholm, J., Tremonti, C., & Leitherer, C. 2018, MNRAS, 481, 1690

Choi, E., Ostriker, J. P., Naab, T., Oser, L., & Moster, B. P. 2015, MNRAS, 449, 4105

Choi, E., Ostriker, J. P., Naab, T., Somerville, R. S., Hirschmann, M., Núñez, A., Hu, C.-Y., & Oser, L. 2017a, ApJ, 844, 31

Choi, J., Conroy, C., & Byler, N. 2017b, ApJ, 838, 159

Chornock, R., Berger, E., Fox, D. B., Lunnan, R., Drout, M. R., Fong, W.-f., Laskar, T., & Roth, K. C. 2013, ApJ, 774, 26

Choudhury, T. R. 2009, Curr. Sci., 97, 841

Choudhury, T. R., Haehnelt, M. G., & Regan, J. 2009, MNRAS, 394, 960

Christensen, C. R., Davé, R., Governato, F., Pontzen, A., Brooks, A., Munshi, F., Quinn, T., & Wadsley, J. 2016, ApJ, 824, 57

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., & Zepf, S. E. 1994, MNRAS, 271, 781

Cole, S., Lacey, C., Baugh, C., & Frenk, C. 2000, MNRAS, 319, 168

Conroy, C. 2013, ARA&A, 51, 393

Conroy, C., & Gunn, J. E. 2010, ApJ, 712, 833

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., White, M., & Gunn, J. E. 2010, ApJ, 708, 58

Cowley, W. I., Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 2018, MNRAS, 474, 2352

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., & Yoshida, N. 2006, MNRAS, 365, 11

Davé, R., Anglés-Alcázar, D., Narayanan, D., Li, Q., Rafieferantsoa, M. H., & Appleby, S. 2019, MNRAS, 486, 2827

Davé, R., Thompson, R., & Hopkins, P. F. 2016, MNRAS, 462, 3265

Dayal, P., & Ferrara, A. 2018, Phys. Rep., 780-782, 1

Dayal, P., Ferrara, A., Dunlop, J. S., & Pacucci, F. 2014, MNRAS, 445, 2545

Dayal, P., Volonteri, M., Choudhury, T. R., Schneider, R., Trebitsch, M., Gnedin, N. Y., Atek, H., Hirschmann, M., & Reines, A. 2020, MNRAS, 495, 3065

De Lucia, G., Kauffmann, G., & White, S. D. M. 2004, MNRAS, 349, 1101

DeBoer, D. R., Parsons, A. R., Aguirre, J. E., et al. 2017, PASP, 129, 045001

Decarli, R., Walter, F., Aravena, M., et al. 2016, ApJ, 833, 70

Dijkstra, M., Gronke, M., & Venkatesan, A. 2016, ApJ, 828, 71

Dijkstra, M., Mesinger, A., & Wyithe, J. S. B. 2011, MNRAS, 414, 2139

Duncan, K., & Conselice, C. J. 2015, MNRAS, 451, 2030

Duncan, K., Conselice, C. J., Mortlock, A., et al. 2014, MNRAS, 444, 2960

Dunlop, J. S., Rogers, A. B., McLure, R. J., et al. 2013, MNRAS, 432, 3520

Dutton, A. A., & van den Bosch, F. C. 2012, MNRAS, 421, 608

Dutton, A. A., van den Bosch, F. C., Faber, S. M., et al. 2010, MNRAS, 410, 1660

Efstathiou, G. 1992, MNRAS, 256, 43P

Eldridge, J. J., Langer, N., & Tout, C. A. 2011, MNRAS, 414, 3501

Eldridge, J. J., & Stanway, E. R. 2009, MNRAS, 400, 1019

—. 2012, MNRAS, 419, 479

Eldridge, J. J., Stanway, E. R., Xiao, L., McClelland, L. A. S., Taylor, G., Ng, M., Greis, S. M. L., & Bray, J. C. 2017, PASA, 34, e058

Ellis, R. S., McLure, R. J., Dunlop, J. S., et al. 2013, ApJ, 763, L7

Emami, N., Siana, B., Alavi, A., Gburek, T., Freeman, W. R., Richard, J., Weisz, D. R., & Stark, D. P. 2019, arXiv:1912.06152

Faber, S. M., & Jackson, R. E. 1976, ApJ, 204, 668

Fan, X., Carilli, C., & Keating, B. 2006a, ARA&A, 44, 415

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., & Fukugita, M. 2006b, AJ, 132, 117

Faucher-Giguère, C.-A., Lidz, A., Hernquist, L., & Zaldarriaga, M. 2008a, ApJ, 682, L9

—. 2008b, ApJ, 688, 85

Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M., & Hernquist, L. 2009, ApJ, 703, 1416

Feng, Y., Di-Matteo, T., Croft, R. A., Bird, S., Battaglia, N., & Wilkins, S. 2016, MNRAS, 455, 2778

Feng, Y., Matteo, T. D., Croft, R., Tenneti, A., Bird, S., Battaglia, N., & Wilkins, S. 2015, ApJ, 808, L17

Finkelstein, S., Dickinson, M., Ferguson, H., et al. 2017, JWST Proposal ID 1345. Cycle 0 Early Release Science

Finkelstein, S. L. 2016, PASA, 33, e037

Finkelstein, S. L., D'Aloisio, A., Paardekooper, J.-P., Russell Ryan Jr., Behroozi, P., Finlator, K., Livermore, R., Sanderbeck, P. R. U., Vecchia, C. D., & Khochfar, S. 2019, ApJ, 879, 36

Finkelstein, S. L., Papovich, C., Giavalisco, M., Reddy, N. A., Ferguson, H. C., Koekemoer, A. M., & Dickinson, M. 2010, ApJ, 719, 1250

Finkelstein, S. L., Papovich, C., Ryan, R. E., et al. 2012a, ApJ, 758, 93

Finkelstein, S. L., Papovich, C., Salmon, B., et al. 2012b, ApJ, 756, 164

Finkelstein, S. L., Ryan, R. E., Papovich, C., et al. 2015a, ApJ, 810, 71

Finkelstein, S. L., Song, M., Behroozi, P., et al. 2015b, ApJ, 814, 95

Finlator, K., Davé, R., & Özel, F. 2011, ApJ, 743, 169

Finlator, K., Keating, L., Oppenheimer, B. D., Davé, R., & Zackrisson, E. 2018, MNRAS, 480, 2628

Finlator, K., Muñoz, J. A., Oppenheimer, B. D., Oh, S. P., Özel, F., & Davé, R. 2013, MNRAS, 436, 1818

Finlator, K., Oh, S. P., Özel, F., & Davé, R. 2012, MNRAS, 427, 2464

Finlator, K., Oppenheimer, B. D., Davé, R., Zackrisson, E., Thompson, R., & Huang, S. 2016, MNRAS, 459, 2299

Finlator, K., Prescott, M. K. M., Oppenheimer, B. D., Davé, R., Zackrisson, E., Livermore, R. C., Finkelstein, S. L., Thompson, R., & Huang, S. 2017, MNRAS, 464, 1633

Finlator, K., Thompson, R., Huang, S., Davé, R., Zackrisson, E., & Oppenheimer, B. D. 2015, MNRAS, 447, 2526

Fioc, M., & Rocca-Volmerange, B. 1997, A&A, 326, 950

—. 1999, arXiv:astro-ph/9912179

—. 2019, A&A, 623, A143

Fletcher, T. J., Tang, M., Robertson, B. E., Nakajima, K., Ellis, R. S., Stark, D. P., & Inoue, A. 2019, ApJ, 878, 87

Foreman-Mackey, D. 2016, J. Open Source Softw., 1, 24

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306

Frebel, A., Johnson, J. L., & Bromm, V. 2009, MNRAS, 392, L50

Frieman, J., Turner, M., & Huterer, D. 2008, ARA&A, 46, 385

Fu, J., Kauffmann, G., Huang, M.-l., Yates, R. M., Moran, S., Heckman, T. M., Davé, R., Guo, Q., & Henriques, B. M. B. 2013, MNRAS, 434, 1531

Fu, J., Kauffmann, G., Li, C., & Guo, Q. 2012, MNRAS, 424, 2701

Furlanetto, S. R., Mirocha, J., Mebane, R. H., & Sun, G. 2017, MNRAS, 472, 1576

Furlanetto, S. R., & Oh, S. P. 2009, ApJ, 701, 94

Furlong, M., Bower, R. G., Theuns, T., et al. 2015, MNRAS, 450, 4486

Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M., & Tremonti, C. A. 2005, MNRAS, 362, 41

Gardner, J. P., Mather, J. C., Clampin, M., et al. 2006, Space Sci. Rev., 123, 485

Geil, P. M., Mutch, S. J., Poole, G. B., Angel, P. W., Duffy, A. R., Mesinger, A., & Wyithe, J. S. B. 2016, MNRAS, 462, 804

Genel, S., Vogelsberger, M., Springel, V., Sijacki, D., Nelson, D., Snyder, G., Rodriguez-Gomez, V., Torrey, P., & Hernquist, L. 2014, MNRAS, 445, 175

Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091

Gilmozzi, R., & Spyromilio, J. 2007, The Messenger, 127, 11

Gnedin, N. Y. 2000, ApJ, 542, 535

—. 2014, ApJ, 793, 29

—. 2016, ApJ, 825, L17

Gnedin, N. Y., Becker, G. D., & Fan, X. 2017, ApJ, 841, 26

Gnedin, N. Y., & Kaurov, A. A. 2014, ApJ, 793, 30

Gnedin, N. Y., & Kravtsov, A. V. 2011, ApJ, 728, 88

Götberg, Y., de Mink, S. E., Groh, J. H., Leitherer, C., & Norman, C. 2019, A&A, 629, A134

Graus, A. S., Bullock, J. S., Boylan-Kolchin, M., & Weisz, D. R. 2016, MNRAS, 456, 477

Grazian, A., Giallongo, E., Paris, D., et al. 2017, A&A, 602, A18

Greif, T. H., & Bromm, V. 2006, MNRAS, 373, 128

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Guaita, L., Pentericci, L., Grazian, A., et al. 2016, A&A, 587, A133

Gunn, J. E., & Peterson, B. A. 1965, ApJ, 142, 1633

Guo, Q., & White, S. D. M. 2009, MNRAS, 396, 39

Haas, M. R., Schaye, J., Booth, C. M., Dalla Vecchia, C., Springel, V., Theuns, T., & Wiersma, R. P. C. 2013, MNRAS, 435, 2931

Harikane, Y., Ouchi, M., Ono, Y., et al. 2016, ApJ, 821, 123

Hassan, S., Davé, R., Finlator, K., & Santos, M. G. 2016, MNRAS, 457, 1550

—. 2017, MNRAS, 468, 122

Hellwing, W. A., Frenk, C. S., Cautun, M., Bose, S., Helly, J., Jenkins, A., Sawala, T., & Cytowski, M. 2016, MNRAS, 457, 3492

Henriques, B. M. B., White, S. D. M., Thomas, P. A., Angulo, R., Guo, Q., Lemson, G., Springel, V., & Overzier, R. 2015, MNRAS, 451, 2663

Hildebrandt, H., Viola, M., Heymans, C., et al. 2017, MNRAS, 465, 1454

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hirschmann, M., Somerville, R. S., Naab, T., & Burkert, A. 2012, MNRAS, 426, 237

Hodge, J. A., Riechers, D., Decarli, R., Walter, F., Carilli, C. L., Daddi, E., & Dannerbauer, H. 2015, ApJ, 798, L18

Hoeft, M., Yepes, G., Gottlöber, S., & Springel, V. 2006, MNRAS, 371, 401

Hopkins, P. F., Kereš, D., Oñorbe, J., Faucher-Giguère, C.-A., Quataert, E., Murray, N., & Bullock, J. S. 2014, MNRAS, 445, 581

Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800

Hubble, E. 1929, Proc. Natl. Acad. Sci. U. S. A., 15, 168

Hui, L., & Gnedin, N. Y. 1997, MNRAS, 292, 27

Iliev, I. T., Mellema, G., Pen, U.-L., Merz, H., Shapiro, P. R., & Alvarez, M. A. 2006a, MNRAS, 369, 1625

—. 2006b, MNRAS, 372, 679

Iliev, I. T., Shapiro, P. R., & Raga, A. C. 2005, MNRAS, 361, 405

Illingworth, G. D., Magee, D., Oesch, P. A., et al. 2013, ApJS, 209, 6

Inoue, A. K. 2003, PASJ, 55, 901

Ishigaki, M., Kawamata, R., Ouchi, M., Oguri, M., Shimasaku, K., & Ono, Y. 2018, ApJ, 854, 73

Izotov, Y. I., Guseva, N. G., Fricke, K. J., Henkel, C., & Schaerer, D. 2017, MNRAS, 467, 4118

Jaacks, J., Finkelstein, S. L., & Bromm, V. 2019, MNRAS, 488, 2202

Jaacks, J., Thompson, R., Finkelstein, S. L., & Bromm, V. 2018, MNRAS, 475, 4396

Jiang, F., & van den Bosch, F. C. 2014, MNRAS, 440, 193

Johns, M. 2008, in Proc. SPIE 6986, Extrem. Large Telesc. Which Wavelengths? Retire. Symp. Arne Ardeb., 698603

Johnson, J. L., Vecchia, C. D., & Khochfar, S. 2013, MNRAS, 428, 1857

Kamionkowski, M., Spergel, D. N., & Sugiyama, N. 1994, ApJ, 426, L57

Katsianis, A., Blanc, G., Lagos, C. P., Tejos, N., Bower, R. G., Alavi, A., Gonzalez, V., Theuns, T., Schaller, M., & Lopez, S. 2017a, MNRAS, 472, 919

Katsianis, A., Tescari, E., Blanc, G., & Sargent, M. 2017b, MNRAS, 464, 4977

Kauffmann, G., White, S. D. M., & Guiderdoni, B. 1993, MNRAS, 264, 201

Kawamata, R., Oguri, M., Ishigaki, M., Shimasaku, K., & Ouchi, M. 2016, ApJ, 819, 114

Kennicutt, Jr., R. C. 1989, ApJ, 344, 685

—. 1998, ApJ, 498, 541

Kennicutt, Jr., R. C., & Evans, II, N. J. 2012, ARA&A, 50, 531

Keres, D., Yun, M. S., & Young, J. S. 2003, ApJ, 582, 659

Kim, H.-S., Wyithe, J. S. B., Raskutti, S., Lacey, C. G., & Helly, J. C. 2013, MNRAS, 428, 2467

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

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

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

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

Klypin, A. A., Trujillo-Gomez, S., & Primack, J. 2011, ApJ, 740, 102

Koekemoer, A. M., Ellis, R. S., McLure, R. J., et al. 2013, ApJS, 209, 3

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330

Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18

Koprowski, M. P., Coppin, K. E. K., Geach, J. E., et al. 2016, ApJ, 828, L21

Kovetz, E. D., Viero, M. P., Lidz, A., et al. 2017, arXiv:1709.09066

Krogager, J.-K., Noterdaeme, P., O'Meara, J. M., et al. 2018, A&A, 619, A142

Kroupa, P. 2001, MNRAS, 322, 231

Krumholz, M. R., & Dekel, A. 2012, ApJ, 753, 16

Krumholz, M. R., McKee, C. F., & Tumlinson, J. 2009, ApJ, 699, 850

Kuhlen, M., & Faucher-Giguère, C.-A. 2012, MNRAS, 423, 862

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

—. 1994, MNRAS, 271, 676

Lacey, C. G., Baugh, C. M., Frenk, C. S., Benson, A. J., Bower, R. G., Cole, S., Gonzalez-Perez, V., Helly, J. C., Lagos, C. D. P., & Mitchell, P. D. 2016, MNRAS, 462, 3854

Lagos, C. d. P., Lacey, C. G., Baugh, C. M., Bower, R. G., & Benson, A. J. 2011, MNRAS, 416, 1566

Lam, D., Bouwens, R. J., Labbé, I., et al. 2019, A&A, 627, A164

LaMassa, S. M., Georgakakis, A., Vivek, M., Salvato, M., Ananna, T. T., Urry, C. M., MacLeod, C., & Ross, N. 2019, ApJ, 876, 50

Laporte, N., Infante, L., Iribarren, P. T., et al. 2016, ApJ, 820, 98

Leitherer, C., Schaerer, D., Goldader, J. D., Delgado, R. M. G., Robert, C., Kune, D. F., de Mello, D. F., Devost, D., & Heckman, T. M. 1999, ApJS, 123, 3

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, ApJ, 737, 12

Leroy, A. K., Walter, F., Brinks, E., Bigiel, F., de Blok, W. J. G., Madore, B., & Thornley, M. D. 2008, AJ, 136, 2782

Levesque, E. M., Leitherer, C., Ekstrom, S., Meynet, G., & Schaerer, D. 2012, ApJ, 751

Liu, C., Mutch, S. J., Angel, P. W., Duffy, A. R., Geil, P. M., Poole, G. B., Mesinger, A., & Wyithe, J. S. B. 2016, MNRAS, 462, 235

Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113

Lo Faro, B., Monaco, P., Vanzella, E., Fontanot, F., Silva, L., & Cristiani, S. 2009, MNRAS, 399, 827

Loeb, A., & Barkana, R. 2001, ARA&A, 39, 19

Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97

LSST Science Collaboration. 2017, arXiv:1708.04058

Lu, Y., Wechsler, R. H., Somerville, R. S., et al. 2014a, ApJ, 795, 123

Lu, Z., Mo, H., Lu, Y., Katz, N., Weinberg, M. D., van den Bosch, F. C., & Yang, X. 2014b, MNRAS, 439, 1294

Ma, X., Hopkins, P. F., Boylan-Kolchin, M., Faucher-Giguère, C.-A., Quataert, E., Feldmann, R., Garrison-Kimmel, S., Hayward, C. C., Kereš, D., & Wetzel, A. 2018a, MNRAS, 477, 219

Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., Zolman, N., Muratov, A. L., Kereš, D., & Quataert, E. 2016a, MNRAS, 456, 2140

Ma, X., Hopkins, P. F., Garrison-Kimmel, S., Faucher-Giguère, C.-A., Quataert, E., Boylan-Kolchin, M., Hayward, C. C., Feldmann, R., & Kereš, D. 2018b, MNRAS, 478, 1694

Ma, X., Hopkins, P. F., Kasen, D., Quataert, E., Faucher-Giguère, C.-A., Kereš, D., Murray, N., & Strom, A. 2016b, MNRAS, 459, 3614

Ma, X., Kasen, D., Hopkins, P. F., et al. 2015, MNRAS, 453, 960

Madau, P. 2017, ApJ, 851, 50

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

Madau, P., Diemand, J., & Kuhlen, M. 2008, ApJ, 679, 1260

Madau, P., Ferguson, H. C., Dickinson, M. E., Giavalisco, M., Steidel, C. C., & Fruchter, A. 1996, MNRAS, 283, 1388

Madau, P., & Fragos, T. 2017, ApJ, 840, 39

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

Maio, U., Ciardi, B., Dolag, K., Tornatore, L., & Khochfar, S. 2010, MNRAS, 407, 1003

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

Maraston, C. 2005, MNRAS, 362, 799

March, M. C., Trotta, R., Berkes, P., Starkman, G. D., & Vaudrevange, P. M. 2011, MNRAS, 418, 2308

Mármol-Queraltó, E., McLure, R. J., Cullen, F., Dunlop, J. S., Fontana, A., & McLeod, D. J. 2016, MNRAS, 460, 3587

Mason, C. A., Fontana, A., Treu, T., et al. 2019, MNRAS, 485, 3947

Mason, C. A., Trenti, M., & Treu, T. 2015, ApJ, 813, 21

Mason, C. A., Treu, T., de Barros, S., Dijkstra, M., Fontana, A., Mesinger, A., Pentericci, L., Trenti, M., & Vanzella, E. 2018a, ApJ, 857, L11

Mason, C. A., Treu, T., Dijkstra, M., Mesinger, A., Trenti, M., Pentericci, L., de Barros, S., & Vanzella, E. 2018b, ApJ, 856, 2

Matthee, J., Sobral, D., Best, P., Khostovan, A. A., Oteo, I., Bouwens, R., & Röttgering, H. 2017, MNRAS, 465, 3637

McClure, R. D., & van den Bergh, S. 1968, AJ, 73, 313

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

McCracken, H. J., Milvang-Jensen, B., Dunlop, J., et al. 2012, A&A, 544, A156

McGreer, I. D., Mesinger, A., & D'Odorico, V. 2015, MNRAS, 447, 499

McGreer, I. D., Mesinger, A., & Fan, X. 2011, MNRAS, 415, 3237

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565

McLeod, D. J., McLure, R. J., & Dunlop, J. S. 2016, MNRAS, 459, 3812

McLeod, D. J., McLure, R. J., Dunlop, J. S., Robertson, B. E., Ellis, R. S., & Targett, T. A. 2015, MNRAS, 450, 3032

McLure, R. J., Cirasuolo, M., Dunlop, J. S., Foucaud, S., & Almaini, O. 2009, MNRAS, 395, 2196

McLure, R. J., Dunlop, J. S., Bowler, R. A. A., et al. 2013, MNRAS, 432, 2696

McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007, MNRAS, 381, 75

McQuinn, M., Lidz, A., Zaldarriaga, M., Hernquist, L., & Dutta, S. 2008, MNRAS, 388, 1101

Mesinger, A. 2010, MNRAS, 407, 1328

Mesinger, A., & Furlanetto, S. 2007, ApJ, 669, 663

Mesinger, A., & Furlanetto, S. R. 2008, MNRAS, 386, 1990

Miralda-Escude, J., Haehnelt, M., & Rees, M. J. 2000, ApJ, 530, 1

Momcheva, I. G., Brammer, G. B., van Dokkum, P. G., et al. 2016, ApJS, 225, 27

Moster, B. P., Naab, T., & White, S. D. M. 2013, MNRAS, 428, 3121

—. 2018, MNRAS, 477, 1822

Moster, B. P., Somerville, R. S., Maulbetsch, C., van den Bosch, F. C., Macciò, A. V., Naab, T., & Oser, L. 2010, ApJ, 710, 903

Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., Hopkins, P. F., Quataert, E., & Murray, N. 2015, MNRAS, 454, 2691

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

Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59

Naidu, R. P., Tacchella, S., Mason, C. A., Bose, S., Oesch, P. A., & Conroy, C. 2020, ApJ, 892, 109

Nakajima, K., Ellis, R. S., Iwata, I., Inoue, A. K., Kusakabe, H., Ouchi, M., & Robertson, B. E. 2016, ApJ, 831, L9

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

Nakajima, K., Fletcher, T., Ellis, R. S., Robertson, B. E., & Iwata, I. 2018, MNRAS, 477, 2098

Narayanan, D., Davé, R., Johnson, B. D., Thompson, R., Conroy, C., & Geach, J. 2018, MNRAS, 474, 1718

Newburgh, L. B., Bandura, K., Bucher, M. A., et al. 2016, Proc. SPIE, 9906, 99065X

Noeske, K. G., Weiner, B. J., Faber, S. M., et al. 2007, ApJ, 660, L43

Noh, Y., & McQuinn, M. 2014, MNRAS, 444, 503

Obreschkow, D., & Rawlings, S. 2009a, MNRAS, 400, 665

—. 2009b, MNRAS, 394, 1857

Ocvirk, P., Aubert, D., Sorce, J. G., et al. 2020, MNRAS, 496, 4087

Ocvirk, P., Gillet, N., Shapiro, P. R., et al. 2016, MNRAS, 463, 1462

Oesch, P. A., Bouwens, R. J., Carollo, C. M., Illingworth, G. D., Trenti, M., Stiavelli, M., Magee, D., Labbé, I., & Franx, M. 2010, ApJ, 709, L21

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., Franx, M., van Dokkum, P. G., Trenti, M., Stiavelli, M., Gonzalez, V., & Magee, D. 2013, ApJ, 773, 75

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., et al. 2014, ApJ, 786, 108

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Labbé, I., & Stefanon, M. 2018, ApJ, 855, 105

Okamoto, T., Gao, L., & Theuns, T. 2008, MNRAS, 390, 920

Oke, J. B., & Gunn, J. E. 1983, ApJ, 266, 713

Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42

Oppenheimer, B. D., & Davé, R. 2006, MNRAS, 373, 1265

O'Shea, B. W., Wise, J. H., Xu, H., & Norman, M. L. 2015, ApJ, 807, L12

Ota, K., Iye, M., Kashikawa, N., et al. 2008, ApJ, 677, 12

Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, ApJ, 723, 869

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

Paardekooper, J.-P., Khochfar, S., & Vecchia, C. D. 2013, MNRAS, 429, 94

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

Padmanabhan, H. 2019, MNRAS, 488, 3014

Parkinson, H., Cole, S., & Helly, J. 2008, MNRAS, 383, 557

Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, MNRAS, 456, 3194

Parsons, A. R., Backer, D. C., Foster, G. S., et al. 2010, AJ, 139, 1468

Patil, A. H., Yatawatta, S., Koopmans, L. V. E., et al. 2017, ApJ, 838, 65

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

Peebles, P. J. 1966, ApJ, 146, 542

Peeples, M. S., Werk, J. K., Tumlinson, J., Oppenheimer, B. D., Prochaska, J. X., Katz, N., & Weinberg, D. H. 2014, ApJ, 786, 54

Pentericci, L., Fontana, A., Vanzella, E., et al. 2011, ApJ, 743, 132

Pentericci, L., Vanzella, E., Fontana, A., et al. 2014, ApJ, 793, 113

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 21

Pillepich, A., Springel, V., Nelson, D., et al. 2018, MNRAS, 473, 4077

Planck Collaboration. 2014, A&A, 571, A16

—. 2016a, A&A, 594, A13

—. 2016b, A&A, 596, A108

—. 2018, arXiv:1807.06209

Poole, G. B., Angel, P. W., Mutch, S. J., Power, C., Duffy, A. R., Geil, P. M., Mesinger, A., & Wyithe, S. B. 2016, MNRAS, 459, 3025

Popping, G., Narayanan, D., Somerville, R. S., Faisst, A. L., & Krumholz, M. R. 2019, MNRAS, 482, 4906

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

Popping, G., Somerville, R. S., & Galametz, M. 2017b, MNRAS, 471, 3152

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

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

Pourtsidou, A. 2018, in Proc. MeerKAT Sci. Pathw. to SKA PoS(MeerKAT2016), Vol. 277 (Trieste, Italy: Sissa Medialab), 037

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

Price-Whelan, A. M., Sipcz, B. M., Günther, H. M., et al. 2018, AJ, 156, 123

Priewe, J., Williams, L. L. R., Liesenborgs, J., Coe, D., & Rodney, S. A. 2017, MNRAS, 465, 1030

Prochaska, J. X., Worseck, G., & O'Meara, J. M. 2009, ApJ, 705, L113

Qin, Y., Mutch, S. J., Poole, G. B., Liu, C., Angel, P. W., Duffy, A. R., Geil, P. M., Mesinger, A., & Wyithe, J. S. B. 2017, MNRAS, 472, 2009

Quinn, T., Katz, N., & Efstathiou, G. 1996, MNRAS, 278, L49

Racca, G. D., Laureijs, R., Stagnaro, L., et al. 2016, Proc. SPIE, 99040O

Rasappu, N., Smit, R., Labbé, I., Bouwens, R. J., Stark, D. P., Ellis, R. S., & Oesch, P. A. 2016, MNRAS, 461, 3886

Rawle, T. D., Egami, E., Bussmann, R. S., et al. 2014, ApJ, 783, 59

Reddy, N. A., Kriek, M., Shapley, A. E., Freeman, W. R., Siana, B., Coil, A. L., Mobasher, B., Price, S. H., Sanders, R. L., & Shivaei, I. 2015, ApJ, 806, 259

Reddy, N. A., Oesch, P. A., Bouwens, R. J., et al. 2018, ApJ, 853, 56

Rémy-Ruyer, A., Madden, S. C., Galliano, F., et al. 2014, A&A, 563, A31

Ricotti, M., & Ostriker, J. P. 2004, MNRAS, 350, 539

Riess, A. G., Macri, L. M., Hoffmann, S. L., et al. 2016, ApJ, 826, 56

Rigby, J., Vieira, J., Bayliss, M., et al. 2017, JWST Propos. ID 1355. Cycle 0 Early Release Sci.

Rigby, J. R., Bayliss, M. B., Sharon, K., Gladders, M. D., Chisholm, J., Dahle, H., Johnson, T., Paterno-Mahler, R., Wuyts, E., & Kelson, D. D. 2018, AJ, 155, 104

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJ, 802, L19

Robertson, B. E., Furlanetto, S. R., Schneider, E., et al. 2013, ApJ, 768, 71

Robitaille, T. P., Tollerud, E. J., Greenfield, P., et al. 2013, A&A, 558, A33

Rodrigues, L. F. S., Vernon, I., & Bower, R. G. 2017, MNRAS, 466, 2418

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

Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, MNRAS, 470, 651

Rogers, A. B., McLure, R. J., Dunlop, J. S., et al. 2014, MNRAS, 440, 3714

Rosdahl, J., Katz, H., Blaizot, J., Kimm, T., Michel-Dansac, L., Garel, T., Haehnelt, M., Ocvirk, P., & Teyssier, R. 2018, MNRAS, 479, 994

Saintonge, A., Lutz, D., Genzel, R., et al. 2013, ApJ, 778, 2

Salpeter, E. E. 1955, ApJ, 121, 161

Sanders, G. H. 2013, J. Astrophys. Astron., 34, 81

Santini, P., Maiolino, R., Magnelli, B., et al. 2014, A&A, 562, A30

Santos, M. G., Cluver, M., Hilton, M., et al. 2017, arXiv:1709.06099

Santos, M. G., Ferramacho, L., Silva, M. B., Amblard, A., & Cooray, A. 2010, MNRAS, 406, 2421

Sawala, T., Frenk, C. S., Fattahi, A., et al. 2016, MNRAS, 456, 85

Schaerer, D. 2003, A&A, 397, 527

Schaerer, D., Izotov, Y. I., Verhamme, A., Orlitová, I., Thuan, T. X., Worseck, G., & Guseva, N. G. 2016, A&A, 591, L8

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schaye, J., Vecchia, C. D., Booth, C. M., Wiersma, R. P. C., Theuns, T., Haas, M. R., Bertone, S., Duffy, A. R., McCarthy, I. G., & van de Voort, F. 2010, MNRAS, 402, 1536

Schechter, P. 1976, ApJ, 203, 297

Schenker, M. A., Ellis, R. S., Konidaris, N. P., & Stark, D. P. 2014, ApJ, 795, 20

Schenker, M. A., Robertson, B. E., Ellis, R. S., et al. 2013, ApJ, 768, 196

Schenker, M. A., Stark, D. P., Ellis, R. S., Robertson, B. E., Dunlop, J. S., McLure, R. J., Kneib, J.-P., & Richard, J. 2012, ApJ, 744, 179

Schinnerer, E., Groves, B., Sargent, M. T., et al. 2016, ApJ, 833, 112

Schmidt, K. B., Treu, T., Bradač, M., et al. 2016, ApJ, 818, 38

Schmidt, K. B., Treu, T., Trenti, M., Bradley, L. D., Kelly, B. C., Oesch, P. A., Holwerda, B. W., Shull, J. M., & Stiavelli, M. 2014, ApJ, 786, 57

Schmidt, M. 1959, ApJ, 129, 243

—. 1963, ApJ, 137, 758

Schroeder, J., Mesinger, A., & Haiman, Z. 2013, MNRAS, 428, 3058

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

Serra, P., Doré, O., & Lagache, G. 2016, ApJ, 833, 153

Shapiro, P. R., Iliev, I. T., & Raga, A. C. 2004, MNRAS, 348, 753

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

Sharma, M., Theuns, T., Frenk, C., Bower, R., Crain, R., Schaller, M., & Schaye, J. 2016, MNRAS, 458, L94

Sharon, C. E., Baker, A. J., Harris, A. I., & Thomson, A. P. 2013, ApJ, 765, 6

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

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

Shibuya, T., Ouchi, M., & Harikane, Y. 2015, ApJS, 219, 15

Shim, H., Chary, R.-R., Dickinson, M., Lin, L., Spinrad, H., Stern, D., & Yan, C.-H. 2011, ApJ, 738, 69

Shivaei, I., Reddy, N. A., Siana, B., et al. 2018, ApJ, 855, 42

Shull, J. M., Harness, A., Trenti, M., & Smith, B. D. 2012, ApJ, 747, 100

Skelton, R. E., Whitaker, K. E., Momcheva, I. G., et al. 2014, ApJS, 214, 24

Smit, R., Bouwens, R. J., Franx, M., Illingworth, G. D., Labbé, I., Oesch, P. A., & van Dokkum, P. G. 2012, ApJ, 756, 14

Smit, R., Bouwens, R. J., Labbé, I., Franx, M., Wilkins, S. M., & Oesch, P. A. 2016, ApJ, 833, 254

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., De Amici, G., Gulkis, S., & Hauser, M. G. 1992, ApJ, 396, L1

Somerville, R. S. 1997, PhD thesis, University of California, Santa Cruz

—. 2002, ApJ, 572, L23

Somerville, R. S., Behroozi, P., Pandya, V., et al. 2018, MNRAS, 473, 2714

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

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

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

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

Somerville, R. S., & Livio, M. 2003, ApJ, 593, 611

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

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

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

Song, M., Finkelstein, S. L., Ashby, M. L. N., et al. 2016, ApJ, 825, 5

Songaila, A., & Cowie, L. L. 2010, ApJ, 721, 1448

Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757

Spergel, D. N., Bean, R., Dore, O., et al. 2007, ApJS, 170, 377

Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175

Springel, V., Yoshida, N., & White, S. D. 2001, New Astron., 6, 79

Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 479, 75

Stanway, E. R., Eldridge, J. J., & Becker, G. D. 2016, MNRAS, 456, 485

Stark, D. P. 2016, ARA&A, 54, 761

Stark, D. P., Ellis, R. S., Charlot, S., et al. 2017, MNRAS, 464, 469

Stark, D. P., Ellis, R. S., Chiu, K., Ouchi, M., & Bunker, A. 2010, MNRAS, 408, 1628

Stark, D. P., Walth, G., Charlot, S., et al. 2015, MNRAS, 454, 1393

Stefanon, M., Bouwens, R. J., Labbé, I., Muzzin, A., Marchesini, D., Oesch, P., & Gonzalez, V. 2017, ApJ, 843, 36

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1

Steidel, C. C., Giavalisco, M., Dickinson, M., & Adelberger, K. L. 1996, AJ, 112, 352

Steidel, C. C., & Hamilton, D. 1992, AJ, 104, 941

—. 1993, AJ, 105, 2017

Sternberg, A., Le Petit, F., Roueff, E., & Le Bourlot, J. 2014, ApJ, 790, 10

Sun, G., & Furlanetto, S. R. 2016, MNRAS, 460, 417

Susa, H., & Umemura, M. 2004a, ApJ, 600, 1

—. 2004b, ApJ, 610, L5

Tacchella, S., Bose, S., Conroy, C., Eisenstein, D. J., & Johnson, B. D. 2018, ApJ, 868, 92

Tacchella, S., Trenti, M., & Carollo, C. M. 2013, ApJ, 768, L37

Tacconi, L. J., Genzel, R., Saintonge, A., et al. 2018, ApJ, 853, 179

Thoul, A. A., & Weinberg, D. H. 1996, ApJ, 465, 608

Tilvi, V., Papovich, C., Finkelstein, S. L., Long, J., Song, M., Dickinson, M., Ferguson, H. C., Koekemoer, A. M., Giavalisco, M., & Mobasher, B. 2014, ApJ, 794, 5

Tilvi, V., Papovich, C., Tran, K.-V. H., et al. 2013, ApJ, 768, 56

Tinker, J., Kravtsov, A. V., Klypin, A., Abazajian, K., Warren, M., Yepes, G., Gottlöber, S., & Holz, D. E. 2008, ApJ, 688, 709

Topping, M. W., & Shull, J. M. 2015, ApJ, 800, 97

Totani, T., Kawai, N., Kosugi, G., Aoki, K., Yamada, T., Iye, M., Ohta, K., & Hattori, T. 2006, PASJ, 58, 485

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

Trac, H., & Gnedin, N. Y. 2011, Adv. Sci. Lett., 4, 228

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

Trebitsch, M., Dubois, Y., Volonteri, M., et al. 2020, arXiv:2002.04045

Trebitsch, M., Volonteri, M., Dubois, Y., & Madau, P. 2018, MNRAS, 478, 5607

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Trenti, M., Perna, R., & Jimenez, R. 2015, ApJ, 802, 103

Trenti, M., Stiavelli, M., Bouwens, R. J., Oesch, P., Shull, J. M., Illingworth, G. D., Bradley, L. D., & Carollo, C. M. 2010, ApJ, 714, L202

Treu, T., Schmidt, K. B., Brammer, G. B., et al. 2015, ApJ, 812, 114

Treu, T., Schmidt, K. B., Trenti, M., Bradley, L. D., & Stiavelli, M. 2013, ApJ, 775, L29

Tully, R. B., & Fisher, J. R. 1977, A&A, 54, 661

van der Burg, R. F. J., Hildebrandt, H., & Erben, T. 2010, A&A, 523, A74

Vanzella, E., de Barros, S., Castellano, M., et al. 2015, A&A, 576, A116

Vanzella, E., Giavalisco, M., Inoue, A. K., et al. 2010, ApJ, 725, 1011

Vanzella, E., Nonino, M., Cupani, G., et al. 2018, MNRAS, 476, L15

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261

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

Visbal, E., & Loeb, A. 2010, JCAP, 2010, 016

Visbal, E., Trac, H., & Loeb, A. 2011, JCAP, 2011, 010

Vogelsberger, M., Genel, S., Sijacki, D., Torrey, P., Springel, V., & Hernquist, L. 2013, MNRAS, 436, 3031

Vogelsberger, M., Genel, S., Springel, V., Torrey, P., Sijacki, D., Xu, D., Snyder, G., Nelson, D., & Hernquist, L. 2014, MNRAS, 444, 1518

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

Wagoner, R. V., Fowler, W. A., & Hoyle, F. 1967in (Cambridge: Cambridge University Press), 1–30

Warren, S. J., Hambly, N. C., Dye, S., et al. 2007, MNRAS, 375, 213

Wechsler, R. H., & Tinker, J. L. 2018, ARA&A, 56, 435

Weisz, D. R., Johnson, B. D., & Conroy, C. 2014, ApJ, 794, L3

Whitaker, K. E., Pope, A., Cybulski, R., Casey, C. M., Popping, G., & Yun, M. S. 2017, ApJ, 850, 208

White, C. E., Somerville, R. S., & Ferguson, H. C. 2015, ApJ, 799, 201

White, S. D. M., & Frenk, C. S. 1991, ApJ, 379, 52

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Wilkins, S. M., Bouwens, R. J., Oesch, P. A., Labbé, I., Sargent, M., Caruana, J., Wardlow, J., & Clay, S. 2016a, MNRAS, 455, 659

Wilkins, S. M., Bunker, A., Coulton, W., Croft, R., Matteo, T. D., Khandai, N., & Feng, Y. 2013, MNRAS, 430, 2885

Wilkins, S. M., Feng, Y., Di Matteo, T., Croft, R., Lovell, C. C., & Waters, D. 2017, MNRAS, 469, 2517

Wilkins, S. M., Feng, Y., Di-Matteo, T., Croft, R., Stanway, E. R., Bouwens, R. J., & Thomas, P. 2016b, MNRAS, 458, L6

Williams, C. C., Curtis-Lake, E., Hainline, K. N., et al. 2018, ApJS, 236, 33

Williams, R. E., Blacker, B., Dickinson, M., et al. 1996, AJ, 112, 1335

Wilson, C. D. 1995, ApJ, 448

Windhorst, R. A., Cohen, S. H., Hathi, N. P., et al. 2011, ApJS, 193, 27

Wise, J. H., Demchenko, V. G., Halicek, M. T., Norman, M. L., Turk, M. J., Abel, T., & Smith, B. D. 2014, MNRAS, 442, 2560

Wise, J. H., Turk, M. J., Norman, M. L., & Abel, T. 2012, ApJ, 745, 50

Wiseman, P., Schady, P., Bolmer, J., Krühler, T., Yates, R. M., Greiner, J., & Fynbo, J. P. U. 2017, A&A, 599, A24

Wong, T., & Blitz, L. 2002, ApJ, 569, 157

Wu, X., Davé, R., Tacchella, S., & Lotz, J. 2020, MNRAS, 494, 5636

Wu, X., McQuinn, M., Kannan, R., D'Aloisio, A., Bird, S., Marinacci, F., Davé, R., & Hernquist, L. 2019, MNRAS, 490, 3177

Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96

Wyithe, J. S. B., Hopkins, a. M., Kistler, M. D., Yüksel, H., & Beacom, J. F. 2010, MNRAS, 401, 2561

Xiao, L., Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 477, 904

Xie, L., De Lucia, G., Hirschmann, M., Fontanot, F., & Zoldan, A. 2017, MNRAS, 469, 968

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

Yoshida, N., Bromm, V., & Hernquist, L. 2004, ApJ, 605, 579

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., Finkelstein, S. L., Popping, G., Davé, R., Venkatesan, A., Behroozi, P., & Ferguson, H. C. 2020a, MNRAS, 496, 4574

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

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

Zahid, H. J., Geller, M. J., Kewley, L. J., Hwang, H. S., Fabricant, D. G., & Kurtz, M. J. 2013, ApJ, 771, L19

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

Zwicky, F. 1937, ApJ, 86, 217