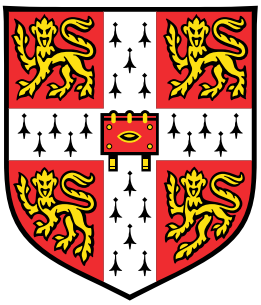
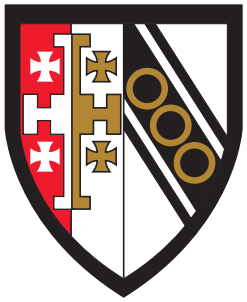
Understanding the Epoch of Reionization Using Simulations of Lyman-α Emitting Galaxies and 21-cm Emission

Lewis Harry Weinberger

Supervisor: Prof. Martin Haehnelt

Institute of Astronomy,
University of Cambridge

This dissertation is submitted for the degree of *Doctor of Philosophy*

Selwyn College June 2020

Abstract

This thesis covers three projects, all centred around using computational simulations to theoretically model the intergalactic medium (IGM) and Lyman-α emitting (LAE) galaxies in the early universe. These galaxies are a key observational probe for understanding the period of the Universe's history known as the Epoch of Reionization (EoR).

In this thesis I explore:

(i) how we can use simulations to model the IGM transmission to LAEs during reionization, and what simulations reveal about the environmental dependence of this transmission. I use these simulations to compare the circumgalactic medium (CGM) and infall environments of bright LAEs residing in more massive host haloes with those of fainter LAEs residing in less massive haloes, and how their environments affect the visibility of LAE populations.

(ii) how to model the population statistics of LAEs and use them to constrain reionization. I implement a model for populating LAEs within dark matter haloes of numerical simulations, which can be used to explore the effect of the IGM on the luminosity function and angular correlation function. By comparing the simulation predictions with current observations it is possible to explore which reionization scenarios are favoured or disfavoured.

(iii) how we can combine LAE and 21-cm observations to derive further constraints on reionization. Using the simulations and modelling established in my first two projects, I explore the cross-correlation of LAEs with 21-cm emission as a further observational statistic for learning about reionization. To compare to observational results I forecast survey sensitivities for a selection of different current and upcoming programs.

Throughout these projects I look in particular at late reionization scenarios, the evidence for which has been growing with recent observational results. By comparing our theoretical late reionization results to those of early reionization scenarios, we can also better understand how to distinguish these possibilities observationally.

Declaration

This thesis is the result of my own work[1] and includes nothing which is the outcome of work done in collaboration, except where specified in the text and in the following chapters:

- Chapter 3, which contains work presented in the paper titled "**Lyman-α emitters gone missing: the different evolution of the bright and faint populations**", published in Monthly Notices of the Royal Astronomical Society, Volume 479, Issue 2, p. 2564-2587. This work was done in collaboration with Girish Kulkarni, Martin G. Haehnelt, Tirthankar R. Choudhury and Ewald Puchwein.
- Chapter 4, which contains work presented in the paper titled "**Modelling the observed luminosity function and clustering evolution of Ly α emitters: growing evidence for late reionization**", published in Monthly Notices of the Royal Astronomical Society, Volume 485, Issue 1, p. 1350-1366. This work was done in collaboration with Martin G. Haehnelt and Girish Kulkarni.
- Chapter 5, which contains work presented in the paper titled "**Probing delayed-end reionization histories with the 21cm-LAE cross-power spectrum**", published in Monthly Notices of the Royal Astronomical Society, Volume 494, Issue 1, p. 703-718. This work was done in collaboration with Girish Kulkarni and Martin G. Haehnelt.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification. It does not exceed the prescribed word limit of 60,000 words.

Lewis Harry Weinberger
June 2020

[1] However throughout the thesis I will use the plural ("*we*"), rather than the singular ("*I*"), to maintain consistency with previously published work. Where necessary I will refer to myself as "*this author*", to distinguish myself (and my contributions to this body of work) from any other collaborator.

I dedicate this thesis to Genny.

Acknowledgements

I would like to thank my supervisor Martin for his unfaltering patience, and for all of the guidance he has given me during this PhD. Whether discussing astrophysics in our weekly meetings, or the tumultuous British political landscape during our group lunches, I have always looked forward to our conversations. His encouragement has kept me on track these past four years, and I have learnt a great deal from him.

I would also like to thank my collaborators for their advice and discussion. In particular I'm very grateful to Girish for showing me the ropes — none of this work would have been possible without his excellent instruction.

I have really enjoyed the environment at the Institute of Astronomy, for which I'd like to thank both my fellow PhD students, the research staff and postdocs, and also the excellent administrative staff. I would like to thank the other budding cryptic crossword solvers for their tolerance in letting me join the daily attempts at tea time, despite the notably mixed results of my Monte Carlo approach to solving the clues. I have also greatly enjoyed the wide range of discussions had over lunch at the exquisite Churchill College cafeteria with colleagues, and during the weekly meetings of the student Cosmology journal club. I would like to thank the past and present members of the numerical simulations group at the IoA for their fellowship; and note that it was a privilege to be able to participate in the DiRAC GPU hackathon as part of team Aregpu.

Finally I would like to acknowledge the amazing support of my family, and my friends outside of the department. My partner Genny has kept me smiling during the ups and downs of the PhD, and has been a constant source of encouragement. My parents have also been unfailing pillars of support, not just through my PhD but at every stage of my life and education. Thank you all, for everything.

Table of contents

List of figures

List of tables

Nomenclature

Notation

a, A	Scalar variable
$\mathbf{a}$, $\mathbf{A}$	Vector variable
$\mathbb{a}$, $\mathbb{A}$	Tensor variable
$\tilde{a}(\mathbf{k})$	Fourier transform of $a(\mathbf{x})$
x_{HI}	Neutral hydrogen fraction
x_{HII}	Ionised hydrogen fraction
Q_{M}	Mass-averaged ionised fraction, also written $\langle x_{\mathrm{HII}} \rangle_{\mathrm{M}}$
Q_{V}	Volume-averaged ionised fraction, also written $\langle x_{\mathrm{HII}} \rangle_{\mathrm{V}}$
Γ_{HI}	Background photoionisation rate
$T_{\mathrm{IGM}}^{\mathrm{Ly}\alpha}$	Lyman-α transmission fraction

Acronyms / Abbreviations

AGN	Active Galactic Nuclei
CGM	Circum-galactic Medium
CMB	Cosmic Microwave Background
DLA	Damped Lyman-α System
EoR	Epoch of Reionization
[R]EW	[Rest-frame] Equivalent Width
GR	General Theory of Relativity
HERA	Hydrogen Epoch of Reionization Array

HSC Hyper Suprime-Cam

IGM Inter-galactic Medium

ISM Inter-stellar Medium

JWST James Webb Space Telescope

LAE Lyman-α Emitting Galaxy

LBG Lyman-break Galaxy

LLS Lyman-limit System

LOFAR Low-Frequency Array

Lyα Lyman-α

Lyβ Lyman-β

LyC Lyman-continuum

MWA Murchison Widefield Array

PFS Prime-Focus Spectrograph

PICO Probe of Inflation and Cosmic Origins

QSO Quasi-stellar Object (Quasar)

RT Radiative Transfer

SKA Square Kilometre Array

SPH Smoothed-Particle Hydrodynamics

UV Ultra-violet

[E/V]LT [Extremely/Very] Large Telescope

WFIRST Wide Field Infrared Survey Telescope (Nancy Grace Roman Space Telescope)

Chapter 1

Introduction

In this introduction we will provide a narrative overview of the history of the universe. This will set the scene for the main time period of interest in this thesis: the Epoch of Reionization (EoR). We will briefly explain the theoretical building blocks required to understand this epoch, and provide examples of recent observational developments.

1.1 Cosmology

In order to introduce the EoR, we must first begin with some preliminary physical cosmology. As theory and observation have often developed hand-in-hand, we will similarly describe them here together.

1.1.1 The Expanding Universe

The force of gravity has played a key role in shaping the universe as we see it today. The General Theory of Relativity (GR) proposed by Einstein (1915) explains this force using a geometric framework. In this theory matter and radiation live in a four dimensional manifold or "*space-time*" which is curved by the presence of said matter and radiation. The motion of objects through space-time is determined by this curvature, giving rise to the familiar force of gravity. Friedmann (1922) first applied this theory in a cosmological context by considering the geometry of a *homogeneous* and *isotropic* universe, with space-time metric given by,

$$\mathrm{d}s^2 = -c^2\mathrm{d}t^2 + a(t)^2\left[\frac{\mathrm{d}r^2}{1-kr^2} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}\phi^2\right)\right], \tag{1.1}$$

where $\mathrm{d}s$ is an infinitesimal line element in the space-time, c is the speed of light, k is a scalar representing the *global* curvature of the universe, $a(t)$ is a scale factor, and (t, r, θ, ϕ) are the space-time coordinates. Further study of this metric and its dynamics was performed by Lemaître (1927); Robertson (1932); Walker (1935), leading to the common moniker: *the FLRW metric*. Astrophysical

observations over the past century such as galaxy mapping (Bengaly et al., 2019) and measurements of the cosmic microwave background (CMB) (Planck Collaboration, 2014) have provided strong evidence that the universe is indeed isotropic about our position. If we assume the Copernican Principle — that we are not unique observers in the universe — then this isotropy also implies homogeneity.

Dynamics of Expansion

One of the properties of an FLRW universe is expansion (or contraction), quantified by the scale factor $a(t)$. The dynamics of this expansion are governed by the Friedmann equations,

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2}, \tag{1.2}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right), \tag{1.3}$$

where $H \equiv \dot{a}/a$ is the Hubble-Lemaitre parameter, ρ is the density, p is the pressure and G is the gravitational constant. On the right hand side of the first Friedmann equation we see that the expansion rate is determined by both the universe's matter (and radiation) content, ρ, and also the curvature k. The expansion of space was confirmed[1] by Hubble (1929), who observed that nearby galaxies were receding from us. More recently, a variety of observations (such as those of Type Ia supernovae in the local universe Riess et al., 1998; Perlmutter et al., 1999) suggest that the universe is not only expanding, but that this expansion is accelerating. The simplest way to incorporate this acceleration into the above Friedmann equations is by adding a constant term, $\Lambda c^2/3$, often referred to as *dark energy*. We will return to the parameters k and Λ in Subsection 1.1.4.

One consequence of this expansion is that radiation travelling through the universe experiences a stretching of its wavelength. This is related to the scale factor as,

$$\frac{\lambda_{\rm obs}}{\lambda_{\rm emit}} = \frac{a(t_{\rm obs})}{a(t_{\rm emit})} \equiv 1 + z, \tag{1.4}$$

where here we have also defined the redshift, z. Throughout the rest of this thesis, redshift will be used as a proxy for cosmological time, and we will use the convention that the scale factor today $a(t_0) \equiv 1$.

We note in passing from Eq. (1.3) that it is possible to have a flat universe with no global curvature ($k = 0$) if the universe has a density equal to,

$$\rho = \frac{3H^2}{8\pi G} \equiv \rho_{\rm crit}. \tag{1.5}$$

[1] Hubble made use of the redshift determinations of Slipher (1917); Strömberg (1925). It should be noted that Lemaître (1927) also used these redshift determinations to calculate the universe's expansion rate, two years prior to Hubble.

1.1.2 The Big Bang

The Friedmann equations admit solutions in which the universe originated from a singular *Big Bang* when $a(t) = 0$. Detailed extrapolation calculations have wound back the clock to determine that our universe did indeed spawn from a hot, low-entropy Big Bang singularity approximately 13.8 billion years ago (Hawking & Ellis, 1973). The second law of thermodynamics suggests that the entropy of an isolated system – which quantifies the number of possible configurations or *degrees of freedom* of a system – should stay the same or increase with time. It might naïvely appear that the hot, uniform 'soup' of primordial matter which existed at the time of the Big Bang would have a larger number of degrees of freedom (and hence constitute a higher entropy state) compared to the colder and more heterogeneous Universe we see today. Importantly this naïve picture neglects the gravitational degrees of freedom of the system; in particular systems in which gravity plays an important role can have a negative heat capacity. Furthermore, the formation of highly non-linear structures such as black holes (Bekenstein, 1974; Hawking, 1975) acts to increase the entropy of the universe. Thus estimates of the entropy of the universe at the time of the Big Bang (dominated by radiation) and today (dominated by black holes), find that the entropy today is as much as a factor $\sim 10^{16}$ times greater (Egan & Lineweaver, 2010).

The Early Universe

As the universe expanded after the Big Bang it also cooled, leading to a number of different "freeze-out" regimes. These regimes all follow a similar pattern: at a given temperature a set of interactions may be possible — for example at very high temperatures in the early universe it was possible for free quarks to exist in thermal equilibrium, interacting via the strong force. However as times passes and the universe cools, eventually a point comes when this set of interactions is no longer energetically favourable, and the interacting species *freeze-out* of equilibrium. This freeze-out occurs when the reaction rate for the interaction, Γ, is overwhelmed by the expansion rate, H,

$$H \gg \Gamma, \tag{1.6}$$

such that the interactions can no longer maintain equilibrium (Weinberg, 2008). To return to the quark example, the free quarks eventually reach low enough energies to be bound up in protons and neutrons.

The early universe proceeded through a number of such stages, passing through regimes where different kinds of interaction dominated and different species of particle dropped out of equilibrium. Two particular regimes of note include,

- The period of Nucleosynthesis (approximately a few seconds/minutes after the Big Bang), when free neutrons and protons cooled to a point where they were able to bind to form nuclei (Alpher et al., 1948). A number of nuclear reactions were possible at these temperatures, leading to the creation of isotopes of mostly hydrogen, helium and lithium. In particular, the most likely

reactions for creating helium and heavier elements involve deuterium, the isotope of hydrogen containing one proton and one neutron,

$$p + n \longrightarrow D + \gamma. \tag{1.7}$$

Deuterium has a binding energy of $\Delta E \approx 2.2$ MeV, however the large of abundance of photons compared to baryons ($\eta = n_b/n_\gamma \approx 10^{-9}$) prevented significant amounts of deuterium from forming until the universe had cooled below $k_\mathrm{B}T \lesssim 0.06$ MeV. This "*deuterium bottleneck*" constrained the abundances of heavier elements produced in this period (Peebles, 1966).

We note that deuterium production is also sensitive to the abundance of protons and neutrons. At early times in this process weak interactions allowed conversion between neutrons and protons (Weinberg, 2008),

$$n + \nu \rightleftharpoons p + e, \tag{1.8}$$

$$n + e^+ \rightleftharpoons p + \overline{\nu}, \tag{1.9}$$

$$n \rightleftharpoons p + e + \overline{\nu}, \tag{1.10}$$

with energy difference $\Delta E = (m_n - m_p)c^2 \approx 1.3$ MeV. As the expansion progressed, eventually the two- and three-body interactions froze out when $k_\mathrm{B}T \approx 0.8$ MeV, leaving only neutron decay to alter the ratio of abundances. Importantly free neutrons have a mean lifetime of $\tau_n \approx 880$ s, resulting in a tight race to form deuterium before the decay of the remaining neutrons (Steigman, 2007). The majority of neutrons which were able to fuse rather than decay then end up in helium; the evolution of the neutron-proton ratio described above results in a mass fraction of primordial helium of $Y_\mathrm{He} \approx 0.24$ (Izotov & Thuan, 1998).

Further nucleosynthesis beyond helium and lithium was inhibited by the non-existence of stable nuclei with 5 or 8 nucleons. We can observe the relative abundances of the primordial elements today (in particular the D/H ratio measured in metal-poor environments), and can use them to infer the baryon density and baryon-photon ratio of the universe (Bludman, 1998).

- The period of Recombination (approximately 400,000 years after the Big Bang), when the universe had cooled sufficiently for the recombination reaction to take place (Peebles, 1968; Zel'dovich et al., 1969),

$$p + e \longrightarrow H + \gamma, \tag{1.11}$$

resulting in the formation of neutral hydrogen. This change in the free electron density also reduced the optical depth to Compton scattering, thus making the universe transparent to radiation. This decoupling period is also referred to as the *time of last scattering*, after which radiation was able to travel unhindered throughout the universe. We observe this relic

radiation today as the Cosmic Microwave Background (CMB), with temperature $T_0 \simeq 2.725$ K. Fluctuations in the temperature of the CMB are a powerful probe of cosmology, as will be discussed further in Subsection 1.1.4.

Although the Big Bang paradigm successfully describes many aspects of the universe that we see today, it does present some problems[2]. Starobinsky (1980); Guth (1981); Linde (1982) suggested that an early period of accelerated expansion, known as *inflation*, would solve a number of these problems as well as offering a number of other appealing features (some of which we will return to in Subsection 1.1.3). Although inflation fits in conveniently with our current theoretical understanding of cosmology, there is currently no *direct* observational evidence for it[3].

1.1.3 Structure Formation

Although in Subsection 1.1.1 we considered an isotropic and homogeneous universe, on small scales the universe we see around us today contains considerable structure such as galaxies and clusters. Observations of the CMB have revealed small temperature perturbations (on the order of $\delta T/\overline{T} \sim 10^{-5}$), some part of which are due to perturbations in the density field away from homogeneity. We will now consider how structure today evolved from these early conditions.

The growth of perturbations

Following recombination the universe was a dark place, filled mostly with neutral hydrogen and the afterglow of the last scattering surface. Density perturbations, which could have been initially seeded by quantum fluctuations (Sakharov, 1966) that were frozen in as classical perturbations during inflation (Hawking, 1982; Bardeen et al., 1983), were able to grow under the force of gravity. We can develop an intuition for this process by considering a simplified scenario: a flat expanding universe filled with non-relativistic matter (with a single particle type of mass m and number density n) in the form of a homogeneous and isotropic fluid. In Newtonian gravity, linear density perturbations to this fluid evolve under the following oscillator equation (Weinberg, 1972),

$$\ddot{\delta} + 2H\dot{\delta} - c_s^2\left(\frac{1}{a^2}\nabla^2 + k_J^2\right)\delta = 0, \tag{1.12}$$

where the perturbation is defined as a density contrast $\delta \equiv \delta n/\overline{n}$, c_s is the sound speed in the fluid, and we have introduced the (physical) Jeans wave-number (Jeans, 1902),

$$k_J^2 = \frac{4\pi G m \overline{n}}{c_s^2}, \tag{1.13}$$

[2] An example being the *horizon problem* (Rindler, 1956): observations of the CMB suggest that the universe was in thermal equilibrium across scales that were causally disconnected in the past, if we naively extrapolate the expansion back in time to the Big Bang.

[3] One possible observational signature of inflation is the creation of "B-mode" polarisation in the CMB (Boyle et al., 2006), however this is non-trivial to disentangle from other sources of B-modes such as dust and lensing.

not to be confused with the curvature parameter k described in the previous section. We can see an interplay between different actors in Eq. (1.12). The second term, referred to as the *Hubble drag*, acts like a frictional force to dampen perturbations. The third dispersion term is key for determining whether a given perturbation mode will grow, decay or oscillate about the background. In particular there are long wavelength modes ($k \ll k_J$) that can grow to form structure. This simplified picture can be expanded to consider how density perturbations will grow in regimes where radiation or the cosmological constant instead dominate the energy density of the universe, and further into a full general-relativistic picture for the non-linear regime (Lifshitz & Khalatnikov, 1964).

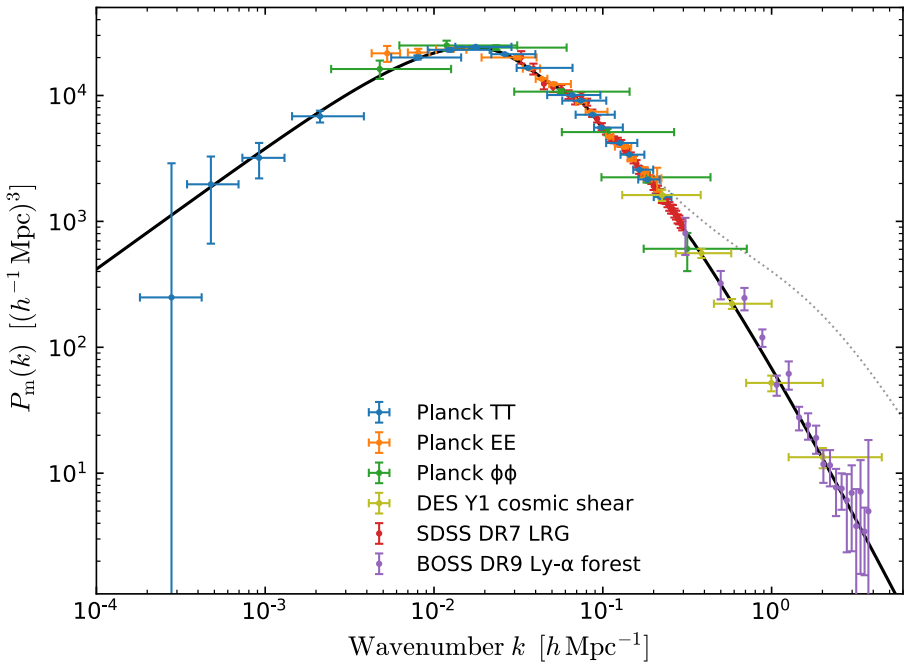


Fig. 1.1 The linear theory matter power spectrum today ($z = 0$) from Planck Collaboration (2018a). Markers indicate observational constraints from the CMB, galaxy distribution, quasar spectra and galaxy weak lensing. The impact of non-linear evolution at $z = 0$ is shown by the grey dotted line. This figure was made by ESA and the Planck Collaboration.

The distribution of fluctuations

We can extend this perturbation picture to predict the statistical distribution of structure (e.g. galaxies) today. The distribution is quantified by the power spectrum $P(k)$ defined as (Peebles, 1980),

$$\langle \tilde{\delta}(\mathbf{k})\tilde{\delta}^*(\mathbf{k}')\rangle = (2\pi)^3 \delta_D(\mathbf{k}-\mathbf{k}')P(k), \tag{1.14}$$

or its Fourier transform the correlation function $\xi(r)$. Note that the variance of fluctuations is related to the power spectrum through,

$$\sigma_\delta^2 = \frac{1}{(2\pi)^3}\int {\rm d}k\, P(k)\, 4\pi\, k^2 = \frac{1}{2\pi^2}\int {\rm d}\ln k\, P(k)\, k^3 \tag{1.15}$$

from which we can also define a *dimensionless power spectrum* which represents the contribution to the variance per logarithmic interval in k,

$$\Delta^2(k) = \frac{k^3 P(k)}{2\pi^2}. \tag{1.16}$$

Although we can calculate the evolution of structure (using for example perturbation theory, or numerical simulations), it still remains to specify the initial conditions. A simple assumption is that the initial perturbations follow a Gaussian distribution[4], and are *adiabatic* in nature such that the perturbations of different components (dark matter, baryonic matter, radiation, etc.) are correlated. Inflationary theories predict primordial density fluctuations with a power-law form,

$$P(k) \propto k^n, \tag{1.17}$$

where n is the spectral index. The scale-invariant[5] case of $n = 1$ is referred to as the *Harrison-Zeldovich* power spectrum (Harrison, 1970; Zeldovich, 1972).

Observations of the matter power spectrum today derived from the CMB and spatial galaxy distribution are in remarkably good agreement with the predictions made using the simple assumptions above, as can be seen in Figure 1.1. This figure shows inferences on the linear matter power spectrum at $z = 0$ (assuming the Planck best-fit ΛCDM model, see Section 1.1.4) from a variety of different observational probes (Chabanier et al., 2019; Tegmark & Zaldarriaga, 2002), including,

- The CMB temperature, E-mode polarisation and lensing reconstruction power spectra from Planck (Planck Collaboration, 2018a).
- Galaxy weak-lensing shear measurements from the Dark Energy Survey (DES) Year 1 results (Troxel et al., 2018). These measurements probe the non-linear matter power spectrum and so require a model to deconvolve non-linear effects, and thus recover the linear matter power spectrum on larger scales than measured.
- Luminous red galaxies (LRGs) from the Sloan Digital Sky Survey (SDSS) data release 7 (Reid et al., 2010). These measurements probe the galaxy power spectrum, so require a model for the galaxy bias to recover the underlying matter power spectrum.
- Lyα forest measurements from SDSS's Baryon Oscillation Spectroscopic Survey (BOSS) data release 9 (Ahn et al., 2012). These measurements probe the 1D flux power spectrum in quasar spectra, and so require modelling a scale and redshift dependent bias to recover the 3D matter power spectrum.

These observations suggest that primordial fluctuations are very close to scale-invariant, with an index of $n \simeq 0.96$ (see Table 1.1).

[4] This also means that we only require the power spectrum to specify the distribution.

[5] The power spectrum of fluctuations in the gravitational potential is given by $P_\Phi(k) \propto k^{-4} P(k)$. The dimensionless version of this is therefore $\Delta^2_\Phi(k) \propto k^{n-1}$, for which $n = 1$ gives Δ^2_Φ independent of k.

Collapse of virial structure

The final stage in the growth of a perturbation is gravitational collapse into virialised structures. The collapse of baryonic matter into stars, black holes and galaxies involves a cosmic dance with its partners, radiation and dark matter. One form of dark matter that has been postulated is the collisionless fluid referred to as *cold dark matter* (Peebles, 1982; Davis et al., 1985). Dark matter does not interact electromagnetically, so it is not supported by thermal pressure against gravitational collapse. This means that at early times dark matter perturbations grow sooner than baryonic perturbations, which are still in thermal equilibrium with radiation (Eisenstein et al., 1998). The dark matter can collapse into virialised haloes, providing a gravitational well into which the baryonic matter can fall and form galaxies. This picture is sometimes referred to as *hierarchical galaxy formation* (White & Rees, 1978). This is a bottom-up picture in which small-scale structures collapse first, and then are incorporated into larger structures; for example the formation of stars inside galaxies, which might themselves form part of clusters. This can be seen more clearly in the *dimensionless* matter power spectrum from Eq. (1.16) which rises with k and then flattens on small scales (note in Figure 1.1, at large k the power spectrum varies as $P(k) \propto k^{n-4}$, such that for $n \simeq 1$ the dimensionless power is approximately constant on these scales).

1.1.4 ΛCDM

The first Friedmann equation (including the Λ constant term) can be re-written,

$$H = H_0 \sqrt{\Omega_{\rm M}(1+z)^3 + \Omega_{\rm k}(1+z)^2 + \Omega_\Lambda}, \tag{1.18}$$

where H_0, $\Omega_{\rm M}$, $\Omega_{\rm k}$ and Ω_Λ are parameters which define the material content of the universe. We note the commonplace definition that $H_0 \equiv 100h$ km/s/Mpc. The Ω density parameters are defined with respect to the critical density of the universe,

$$\Omega_X \equiv \frac{\rho_X}{\rho_{\rm crit}}. \tag{1.19}$$

Ingredients such as these parameters can be used to construct models that aim to explain all of the observed behaviour of the universe. There is now consensus that the 6-parameter "concordance" ΛCDM model is able to satisfactorily explain many key cosmological observations. Inference on this model with data from the CMB (Planck Collaboration, 2018b; Hinshaw et al., 2013), baryonic acoustic oscillations (BAO) (Alam et al., 2017), and weak lensing (Abbott et al., 2018) finds mostly consistent parameters[6]. In Table 1.1 we show 6 independent parameter values inferred from observations of the CMB with the Planck telescope (Planck Collaboration et al., 2014; Planck Collaboration, 2018b). We also show derived parameters which are relevant to this work.

[6]Some *tensions* have been noted recently, in particular regarding "early-time" (such as Riess et al., 2019) and "late-time" (such as Planck Collaboration, 2018b) measurements of the Hubble constant H_0.

Table 1.1 68% confidence limits on best fit parameters constrained by Planck Collaboration et al. (2014); Planck Collaboration (2018b).

Parameter	Description	Planck 2013	Planck 2018
$\Omega_b h^2$	Baryon density today	0.02205 ± 0.00028	0.0224 ± 0.0001
$\Omega_c h^2$	(Cold) dark matter density today	0.1199 ± 0.0027	0.120 ± 0.001
$100\theta_*$	Angular scale of the sound horizon at last scattering	1.04147 ± 0.00062	1.0410 ± 0.0003
τ	Thomson optical depth	$0.089^{+0.012}_{-0.014}$	0.054 ± 0.007
n_s	Scalar spectral index	0.9603 ± 0.0073	0.965 ± 0.004
$\ln\left(10^{10}A_s\right)$	Amplitude of curvature power spectrum	$3.089^{+0.024}_{-0.027}$	3.044 ± 0.014
H_0	Expansion rate (in km/s/Mpc)	67.3 ± 1.2	67.4 ± 0.5
Ω_m	Total matter density today	0.315 ± 0.017	0.315 ± 0.007
Ω_Λ	Dark energy density today	0.685 ± 0.017	0.685 ± 0.007
σ_8	Root-mean-square matter fluctuations today	0.829 ± 0.012	0.811 ± 0.006
$Y_{\rm He}$	Primordial baryonic fraction in Helium	0.24770 ± 0.00012	0.2454

The ΛCDM model is a quantification of the cosmological paradigm that has been described in the previous sections. This paradigm suggests that,

- our universe is very close to being flat, with $k \simeq 0$ suggested by joint analysis of observations of the CMB with BAO (Efstathiou & Gratton, 2020);
- there is a dark energy component given by the cosmological constant, Λ, which makes up roughly 70% of the energy density of the universe and causes an acceleration in its expansion (Riess et al., 1998; Perlmutter et al., 1999);
- there is a cold dark matter component, which makes up roughly 85% of the total matter density (25% of the total energy density of the universe) and is essential for the growth of structure (Blumenthal et al., 1984).

Whilst an effective parametrisation and paradigm, it is important to highlight that the model gives no explanation for what the dark components fundamentally are. Furthermore, a number of problems have been suggested to result in this framework such as the *core-cusp* problem of dark matter (Navarro et al., 1996) in which simulated low mass haloes are more concentrated than their observed counterparts, dwarf galaxies. We note also that there are many alternative models to ΛCDM[7] although at present none are observationally favoured over ΛCDM.

[7]Examples include quintessence (Caldwell et al., 1998), in which Λ is not constant but a function of time, and warm dark matter (Colombi et al., 1996) which behaves somewhat differently to cold dark matter on small scales.

1.2 The Epoch of Reionization

Having established an outline of physical cosmology and the history of the early universe, we can now consider the era of Reionization. We will sketch the timeline for this period, including the physical processes and players involved, before considering the latest observational constraints.

1.2.1 A Brief History

The term *reionization* refers to the ionisation of primordial hydrogen and helium, the majority of which was frozen out in a neutral state following the Recombination era. The energy required to ionise an H I atom[8] to H II is 13.6 eV, whilst He I requires 24.6 eV for first ionisation to He II, and a further 54.4 eV to become He III. The Epoch of Reionization therefore encompasses *three* different ionisation transitions, each of which became possible as ionisation sources start radiating at the required energies (Meiksin, 2009). This effectively reduces to two distinct periods in the universe's history: (i) H I reionization (and at approximately the same time He I, due to the similar ionisation energy), and (ii) He II reionization. The primary concern of this thesis work is H I reionization.

H I reionization (hereafter simply reionization) can itself be divided into a number of distinct stages (Loeb & Barkana, 2001). As discussed in Subsection 1.1.3, galaxies form in dark matter haloes which reside in over-dense regions. The first stage of reionization, "pre-overlap", involves galaxies forming and beginning to leak ionising radiation into the surrounding inter-galactic medium (IGM) (Gnedin, 2000a). When the surrounding over-density has been ionised, the ionisation front can rapidly pass through any nearby under-dense regions until they meet other ionisation fronts. This pre-overlap phase is highly inhomogeneous (Ciardi et al., 2000), with the local ionisation field dependent on the presence of nearby sources and the small-scale density distribution. The meeting of different ionisation fronts marks the beginning of a relatively fast "overlap" stage, where ionising photons from multiple sources are present within ionised regions. This increase in the intensity of radiation can rapidly ionise the low-density IGM, all except very over-dense pockets which are able to "self-shield" and remain neutral (Miralda-Escudé et al., 2000; Chardin et al., 2018). When the ionised bubbles have fully percolated, the ionising radiation starts to form a more uniform background. The final "post-overlap" stage of reionization progresses indefinitely, as collapsed objects continue to form H I but the rest of the low-density IGM has been ionised (Hernquist et al., 1996).

The formation of cosmological H II regions in the IGM is somewhat different to the H II regions which form in the inter-stellar medium (ISM) around stars. An isotropic H II region expands into the IGM according to (Shapiro & Giroux, 1987; Shapiro, 1986),

$$4\pi R_{\rm ion}^2 \left(\frac{{\rm d}R_{\rm ion}}{{\rm d}t} - H R_{\rm ion} \right) = \frac{\dot{N}_{\rm ion}}{\langle n_{\rm H} \rangle} - \frac{4}{3} \frac{\pi R_{\rm ion}^3}{t_{\rm rec}} \tag{1.20}$$

[8] We use the astrophysical notation for ionisation state, where a neutral atom has suffix I, and higher ionisation states increment this roman numeral (II, III, etc.).

where $R_{\rm ion}$ is the proper radius of the ionisation front, $\langle n_{\rm H} \rangle$ is the proper (spatial) average number density of H atoms, $\dot{N}_{\rm ion}$ is the ionising emission rate from the central source, and the recombination time is given by,

$$t_{\rm rec} = \frac{1}{\langle n_{\rm H} \rangle f_e \alpha_B(T) C}, \tag{1.21}$$

where $C = \left(\langle n_{\rm H}^2 \rangle / \langle n_{\rm H} \rangle^2 \right)^{1/2}$ is the clumping factor, f_e is the number of electrons per H atom, and α_B is the case B recombination coefficient. The expansion of the universe seen in the second term of Eq. (1.20) reduces the density of surrounding H I gas as the ionisation front passes, thus preventing the formation of a Strömgren sphere[9] (Meiksin & Madau, 1993).

The average progress of reionization, quantified by the ionisation volume-filling factor[10] $Q_{\rm V} = \langle x_{\rm HII} \rangle_V$ is governed by (Madau et al., 1999),

$$\frac{{\rm d}Q_{\rm V}}{{\rm d}t} = \frac{\dot{n}_{\rm ion}}{\langle n_{\rm H} \rangle} - \frac{Q_{\rm V}}{t_{\rm rec}} \tag{1.22}$$

where $\dot{n}_{\rm ion}$ is the ionising emission rate per comoving volume. This average evolution is essentially an exercise in photon counting: balancing ionisations and recombinations. When ionising sources are producing more photons than there are hydrogen atoms, and at a rate greater than the number of recombinations, the ionisation volume-filling factor grows. As seen in Eq. (1.21), the recombination time increases as the proper hydrogen density decreases, which evolves with the expansion as $\langle n_{\rm H} \rangle \propto (1+z)^3$.

Early ionising sources

The evolution of reionization is determined by the distributions of the H I gas and the ionising sources. The H I distribution is determined by the formation of large-scale structure, however the nature of the ionising sources responsible for reionization is less well known (Couchman & Rees, 1986; Mirabel et al., 2011; Dopita et al., 2011; Katz & Ricotti, 2014; Madau & Haardt, 2015; Sharma et al., 2016).

During the dark ages that followed recombination, the first stars in the universe collapsed into existence, collectively forming the first galaxies. As star formation is key for the production of metals[11], this first population — otherwise known as Population III stars (Bromm & Larson, 2004) — formed in a metal-free environment. The lack of metals in Population III stars results in a very different evolution compared to the enriched stars we observe today, known as Population I & II stars (McKee & Tan, 2008). In particular this lack of metals removes many common cooling mechanisms, thus allowing these early stars to reach large masses $M \sim 100 M_\odot$ (Abel et al., 2002). Theoretical calculations predict that alongside large masses and temperatures, Population III stars would also

[9] Stellar H I regions eventually reach a point of equilibrium where the rate of ionisations is balanced by the rate of recombinations within some volume (Strömgren, 1932).

[10] Also referred to as the *porosity* parameter for $Q < 1$. An alternative quantity to consider is the *mass-weighted* ionisation fraction (Chen et al., 2020).

[11] Metallicity is a measure of the abundance of elements heavier than hydrogen or helium.

have had very high ionising photon outputs (Tumlinson, 2000), potentially making them ideal sources for the early[12] stages of reionization. Observations of high-redshift galaxies (Robertson et al., 2013) allow us to extrapolate back in time to try to determine their contribution to the ionising photon budget. A key unknown in this determination is the escape fraction, $f_{\rm esc}$, of photons from within a galaxy. The photons produced in stars need to escape through their local ISM in order to propagate out in the IGM (Wood & Loeb, 2000). Both numerical simulations such as Katz et al. (2019b), and observations of local analogs such as Green Pea galaxies (Yang et al., 2016), suggest that the galaxy population should have been sufficient to reionize the IGM if early galaxies had escape fractions on the order of $f_{\rm esc} \sim 10 - 20\%$ (Haardt & Madau, 2012).

Another possible candidate for providing the needed ionising photons is the bright active galactic nuclei (AGN) at the centres of galaxies (Madau & Haardt, 2015; Chardin et al., 2017). We observe these super-massive black holes in the form of quasi-stellar objects (QSOs, or quasars) in both the local universe (Schmidt, 1963) and also at high redshifts (Bañados et al., 2018). Unlike stellar radiation within galaxies, the ionising output from AGN is concentrated in a single central source that can easily penetrate into the surrounding IGM (Miralda-Escudé, 2003). However extrapolations of the number density of quasars in the early universe suggest that their emissivity is too low to have caused reionization unless there is a population of unobserved, low-luminosity quasars (Kulkarni et al., 2019b). Another problem with a quasar-driven reionization is that it would require a hard ionising spectrum, which would also lead to an earlier He II reionization than is observed (Mitra et al., 2018). It is also possible, however, that a population of mini-quasars (Madau et al., 2004; Dijkstra et al., 2004) at $z > 6$ could provide some contribution to reionization, if intermediate-mass black holes (IMBH) are able to accrete as quasars.

Other, more exotic, ionising sources could be present in the early universe, contributing to reionization. These include an early generation of stellar black holes (Mirabel et al., 2011), and globular clusters (Katz & Ricotti, 2014).

An example theoretical reionization history is shown in Figure 1.2. In this model, from Puchwein et al. (2019), reionization is driven by galaxies which leak $f_{\rm esc} \lesssim 18\%$ of their ionising radiation, whilst the later He II reionization is driven by AGN.

Temperature evolution and feedback

Alongside driving the phase transition from H I to H II, the ionising photons also played a role in photo-heating the gas in the IGM. Observations of quasar spectra today suggest that the IGM was heated to temperatures of approximately $T \sim 10^4 K$ by the end of reionization (Hui & Gnedin, 1997; Haehnelt & Steinmetz, 1998; McQuinn & Upton Sanderbeck, 2016). Ionised hydrogen can radiatively cool through recombinations, bremsstrahlung, and inverse Compton scattering with CMB photons. By the end of reionization the low density IGM settled into a power-law relationship

[12] We note that the short-lived Population III stars enrich their surrounding ISM, such that a new generation of more metal-rich stars can begin to form.

between temperature and density, which holds over two orders of magnitude in density,

$$T = T_0 \Delta^{\gamma - 1}, \tag{1.23}$$

where T_0 is the temperature at mean density, Δ is the baryonic density relative to the cosmic mean, and γ is the adiabatic index (sometimes referred to as the *equation of state*). The latest constraints from Lyman-α forest measurements by Gaikwad et al. (2020) find that $T_0 = 120000 \pm 2200$ K and $\gamma = 1.05 \pm 0.22$ at $z \sim 5.8$, however there is little evolution between $5.3 < z < 5.9$.

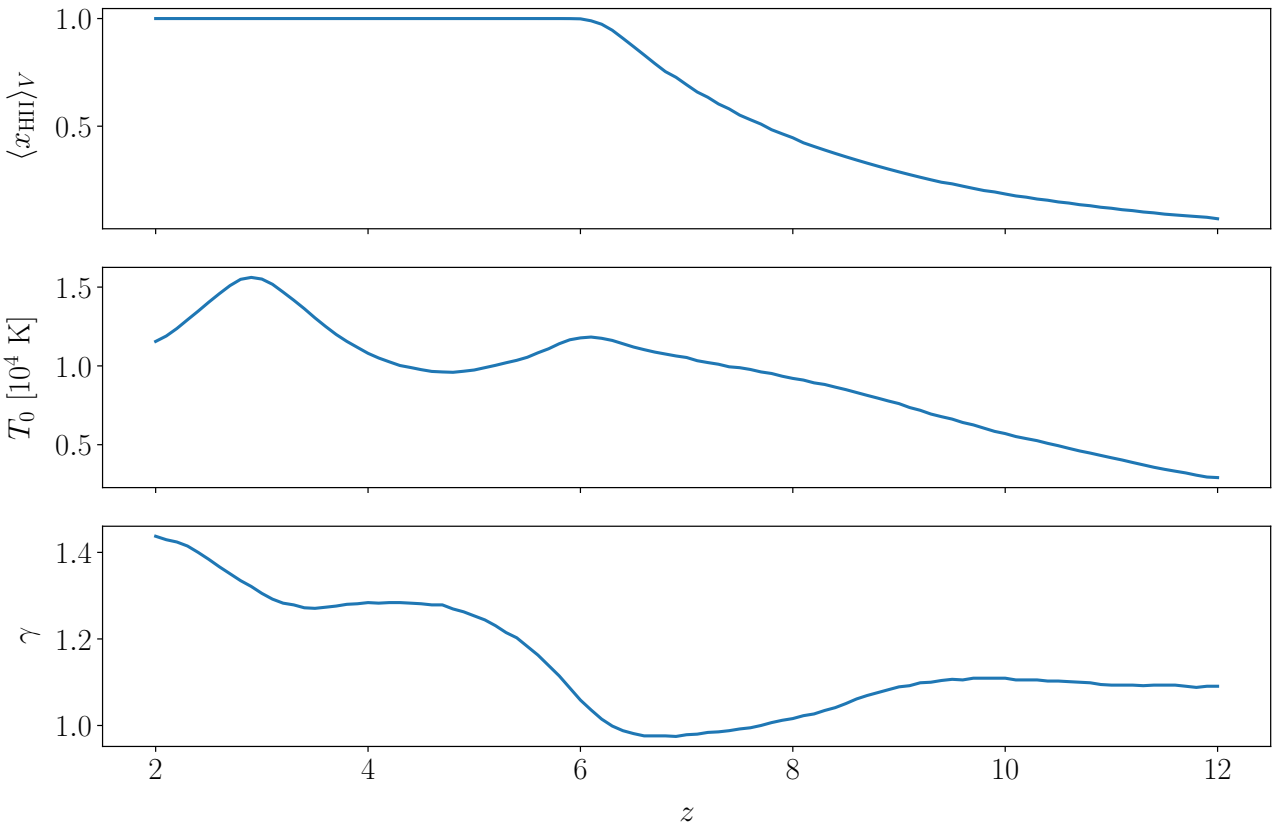


Fig. 1.2 An example reionization history as a function of redshift, from Puchwein et al. (2019). *Top*: the volume-filling factor, which evolves from a fully neutral IGM ($x_{\rm HII} = 0$) to an ionised IGM ($x_{\rm HII} = 1$) by $z \sim 6$. *Middle*: the temperature of the IGM at mean density, driven by photo-heating during reionization. *Bottom*: the adiabatic index which controls the IGM equation of state.

The presence of an ionising UV background within ionised regions can also impact the star formation processes in galaxies (Gnedin, 2000b). The mass threshold for a proto-stellar cloud to collapse into a star is a strong function of temperature, as was seen in Eq. (1.13) via the sound speed c_s. On larger scales this means that in the photo-heated environment of reionization, gas infall onto less massive haloes will be disrupted — impacting the formation of less massive galaxies (Dawoodbhoy et al., 2018; Katz et al., 2019a).

1.2.2 Observational Probes and Constraints

Observing reionization directly presents a number of challenges. As a high-redshift ($z \gtrsim 5$) phenomenon, there are a number of confounding obstacles that make observing galaxies difficult: large proper distances to sources, smaller populations at an earlier stage of evolution and the redshift of their emission to longer wavelengths[13] (Alvarez et al., 2019). A few methods have been used to overcome these difficulties in order to constrain the progress of reionization, and indirectly infer the state of the IGM, at high redshifts.

Our constraints and understanding of how — and when — exactly reionization took place have evolved over the past decades as observations have expanded the frontiers of astronomical knowledge. For example at the turn of the century, Haehnelt et al. (2001) interpreted Lyman-break galaxy (LBG) observations to suggest that reionization was driven by massive stars within early galaxies, whilst WMAP observations of the CMB (Kogut et al., 2003; Spergel et al., 2003) implied that reionization took place at very high redshifts around $z \sim 17$. A decade later, the initial release of results from the Planck satellite (Planck Collaboration et al., 2014; Planck Collaboration, 2016) increased interest in more intermediate-redshift reionization scenarios ($z \sim 9$), whilst new observations of the quasar luminosity function (Giallongo et al., 2015) made quasar-driven reionization more favourable (Madau & Haardt, 2015). There is now growing evidence from the final Planck results (Planck Collaboration, 2018b), as well as measurements of the Lyman-α forest (Becker et al., 2018) and Lyman-α emitting galaxies (Konno et al., 2018) suggesting that reionization may have been considerably *later* than first thought, ending as late as $z \sim 5.3$ (Kulkarni et al., 2019a). Whilst their escape fraction at high-redshift remains uncertain, galaxies remain the most plausible source of reionizing photons (Finkelstein et al., 2019). Future 21-cm observations may fill in the final missing details in this picture (Koopmans et al., 2015).

The Cosmic Microwave Background

We observe small temperature fluctuations in the CMB from early times, referred to hereafter as the primary anisotropies, which arose due to perturbations in density, velocity and the gravitational potential. Reionization also impacts the CMB radiation that we observe: it generates additional polarisation, causes a damping to the primary anisotropies, and also introduces secondary anisotropies (Haiman & Knox, 1999). The process of reionization increases the density of free electrons which can scatter CMB photons, and thus smear out anisotropies (Efstathiou, 1988). This scattering damps the amplitude of the CMB temperature and polarisation power spectra on small scales by a factor $e^{-2\tau_e}$ where the Thomson optical depth to electron scattering, τ_e, is given by,

$$\tau_e(z) = \int_{t(z)}^{t_0} \mathrm{d}t' \, n_e(t') \, c \, \sigma_T, \tag{1.24}$$

[13] For example Lyman-α emission (in the UV) at $z = 7$ is observed redshifted into the infrared. The James Webb Space Telescope (JWST) will have instruments for observing in the near- and mid-infrared, opening up a new window into high-redshift galaxies.

where σ_T is the cross-section for scattering, and the number density of free electrons at a given time, $n_e(t)$, is determined by reionization. Although this suppression can be significant, it is degenerate with the amplitude of the primordial power spectrum, A_s. However, the increase in the free electron density during reionization has a further effect on the CMB: radiation with a quadrupole anisotropy that Thomson scatters off these electrons can generate linear polarisation (Rees, 1968). When decomposed into the E- and B-mode basis, this reionization-generated polarisation produces a distinctive peak in the EE power spectrum on scales larger than the horizon size at reionization, with an amplitude that scales $\propto \tau_e^2$. Fortunately the other parameters of ΛCDM cannot produce such a peak, and thus this "reionization bump" is a powerful probe of τ_e with minimal degeneracies (Reichardt, 2016).

The latest Planck measurement by Planck Collaboration (2018b) found a value of $\tau_e = 0.054 \pm 0.007$, considerably lower than earlier measured values from WMAP[14] of $\tau_e = 0.081 \pm 0.012$ (Hinshaw et al., 2013). If a relatively simple evolution is assumed[15] for the ionised fraction $\langle x_{\rm HII}(z) \rangle$, then this implies that the midpoint of reionization was $z_{\rm mid} = 7.64 \pm 0.74$. Note this is considerably later than the estimate from WMAP which suggested instead that reionization occurred closer to $z_{\rm mid} = 10.1 \pm 1.0$.

The secondary anisotropies introduced by reionization arise from two mechanisms: photons scattering off free electrons that are hotter than the CMB (Zeldovich & Sunyaev, 1969), and photons scattering off electrons that are moving with a large-scale bulk velocity (Ostriker & Vishniac, 1986)[16]. These anisotropies are smaller than the primary anisotropies, and more difficult to disentangle, requiring high signal-to-noise observations of the CMB (Trombetti & Burigana, 2018). We will return to the future observational outlook in Chapter 6.

The Lyman-α forest

It is possible to indirectly observe the IGM itself via absorption features in bright background sources. The prototypical example of this is the series of absorption features seen in distant quasar spectra, referred to as the *Lyman-α forest*. Atomic hydrogen, despite being the simplest atomic element, has a number of distinct electronic transitions which can interact radiatively to absorb and produce photons. The transition from $2p \longrightarrow 1s$ is the Lyman-α transition, with an energy difference of 10.2 eV corresponding to a wavelength of 1216 Å (Lyman, 1906), as shown in Figure 1.3. Photons close to this wavelength travelling from a background quasar will be resonantly scattered by any intervening neutral hydrogen, creating an absorption trough in the spectrum (Gunn & Peterson, 1965). Quasars emit at a continuum of wavelengths which will redshift with the expansion; if photons from shorter wavelengths have redshifted close to Lyman-α line-centre by the time they reach pockets of neutral hydrogen, then they will be scattered out of the line of sight. This creates the "forest" structure of troughs in the spectra of quasars observed at intermediate redshifts $z \sim 3$–5 (McDonald et al., 2006).

[14] Prior results from WMAP of $\tau_e = 0.17 \pm 0.06$ (Kogut et al., 2003; Spergel et al., 2003) were measured using the temperature-polarisation (TE) cross-power spectrum rather than the EE auto-power spectrum, and thus had considerably larger errors.

[15] For example assuming the free electron fraction varies as a hyperbolic tangent function of redshift (Lewis, 2008).

[16] These phenomena are known as the *thermal* and *kinetic* Sunyaev-Zeldovich (SZ) effects, respectively.

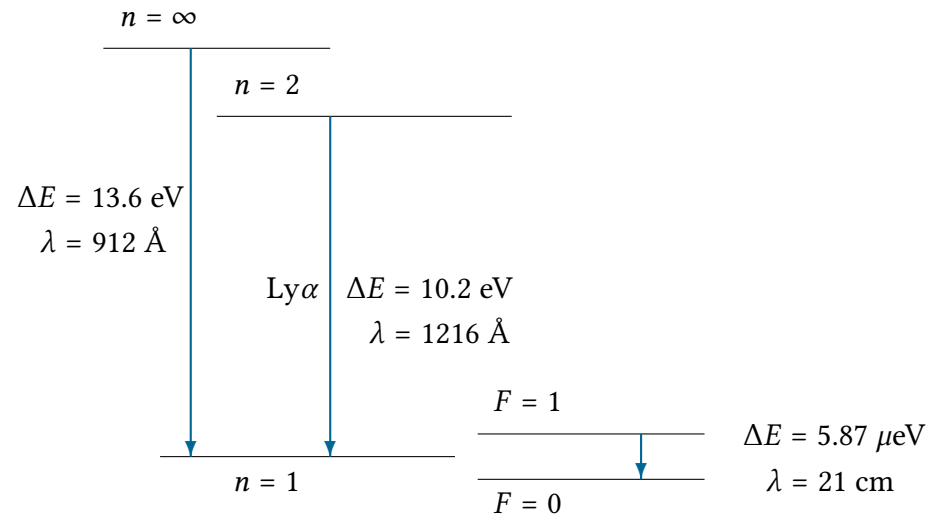


Fig. 1.3 A subset of the electronic structure of the hydrogen atom. The "Lyman-α" line results from the $2p \longrightarrow 1s$ transition, which has a corresponding energy of 10.2 eV. The $1s$ ground state also undergoes a hyperfine splitting into two distinct energy levels due to interactions between the magnetic field generated by the electron and the nuclear dipole moment, giving rise to the "21-cm" line which has an energy of 5.87 μeV. These two lines are key to many astrophysical observations. Also shown is the Lyman-limit transition between $n = 0$ and $n = \infty$, with energy 13.6 eV, which is the energy required to ionise hydrogen from the ground state.

The optical depth of the IGM to scattering radiation via the Lyman-α transition is a function of the radiation frequency. The total optical depth along a sightline to the redshift of emission, $z_{\rm emit}$, measured at an observed frequency, $\nu_{\rm obs}$, is given by the integral,

$$\tau(\nu_{\rm obs}) = \int_0^{z_{\rm emit}} {\rm d}z \, \frac{{\rm d}\ell}{{\rm d}z} \, n_{\rm HI}(z) \, \sigma(\nu_{\rm obs}(1+z)), \tag{1.25}$$

where ${\rm d}\ell/{\rm d}z$ is the proper line element, and the form of the cross-section for a scattering event, $\sigma(\nu) = \sigma_0 \phi(\nu)$, arises from a combination of atomic and thermal physics. The scattering cross-section is strongly peaked around $\nu = \nu_{{\rm Ly}\alpha}$, so we can derive the strength of this so-called "Gunn-Peterson" absorption at redshift z as a function of $x_{\rm HI}$ by approximating the cross-section as a δ-function centred on $\nu_{{\rm Ly}\alpha}$ (Gunn & Peterson, 1965),

$$\tau_{\rm GP}(z) \simeq 4.6 \times 10^5 \, x_{\rm HI} \, (1+z)^{\frac{3}{2}} \, (1 - Y_{\rm He}) \, \frac{\Omega_b h^2}{(\Omega_m h^2)^{1/2}}, \tag{1.26}$$

where the cosmological parameters are as defined in Table 1.1.

In reality the cross section is not a δ-function; instead, as a result of the finite lifetime of an excited quantum state, the discrete frequency of the transition is broadened and the line-shape is

well modelled by a Lorentzian function,

$$\phi_L(\nu) = \frac{1}{\pi} \frac{\Gamma}{(\nu - \nu_{21})^2 + \Gamma^2}, \tag{1.27}$$

where ν_{21} is the Lyman-α frequency, $\Gamma = A_{21}/4\pi$ and $A_{21} = 6.25 \times 10^8 \mathrm{s}^{-1}$ is the radiative decay rate. The thermal motions of the absorbing gas (i.e. hydrogen in the IGM) introduce a further broadening to the cross-section. Assuming the thermal distribution is Maxwellian, the resulting cross-section is a convolution of the Lorentz profile with the Maxwell-Boltzmann probability distribution, $P(v)$, which is called the Voigt function (Tepper-García, 2006),

$$\phi(\nu) = \int \mathrm{d}v \, P(v) \phi_L \left(\nu - \frac{\nu v}{c} \right). \tag{1.28}$$

From Eq. (1.27) we can see that the probability of a scattering event via the Lyman-α transition is therefore very sensitive to how close the radiation frequency is to the Lyman-α frequency, giving rise to the term *resonant scattering*. We note also that the Lorentz part of the Voigt profile introduces expanded wings around the core part of the profile, which is Gaussian.

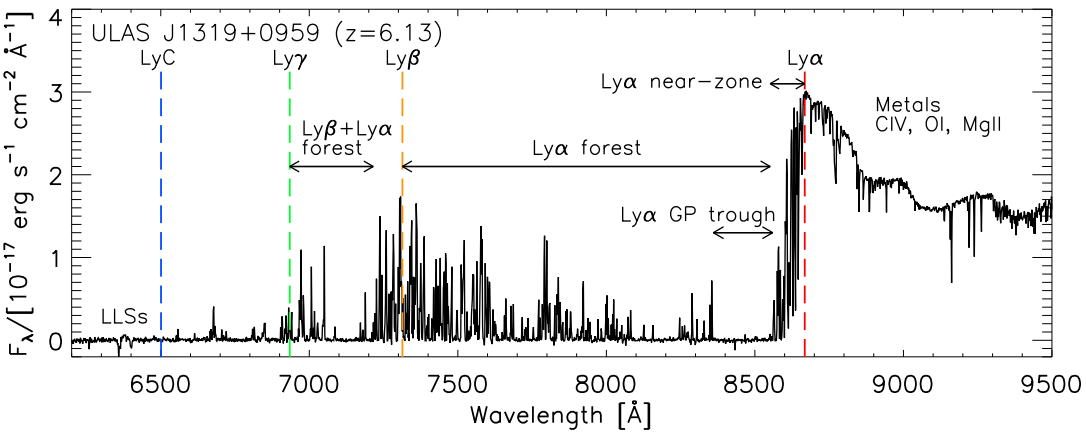


Fig. 1.4 A schematic $z = 6.13$ quasar spectrum from Becker et al. (2015a) showing the Lyman-α forest features. This spectrum was obtained by Becker et al. (2015b) using the X-Shooter telescope. This quasar is sufficiently high redshift to show significant troughs where neutral pockets of gas along the sight-line are scattering the background light out of the sightline.

As seen in Eq. (1.26), absorption of Lyman-α radiation is also very sensitive to the density of neutral hydrogen. This makes the Lyman-α forest a useful probe of the *end* of reionization, where most of the hydrogen in the IGM has been ionised, leaving only pockets of neutral hydrogen in over-densities and collapsed structures. At earlier times before the end of reionization, there is sufficient neutral hydrogen in the intervening IGM that the forest is no longer observed in high redshift quasars, but instead there is a single absorption trough bluewards of the Lyman-α wavelength.

Fan et al. (2006) observed this phenomenon to determine that the end of reionization was close to $z \sim 6$.

Recent spectroscopic surveys of high redshift quasars (Becker et al., 2018, 2015b) have identified a large sightline-to-sightline scatter in the average optical depth. This includes quasars at the same redshift that have spectra with considerable transmission, and others with extensive absorption[17]. Theoretical modelling of reionization has struggled to reproduce this scatter, with suggested mechanisms including fluctuations in the temperature, background photoionisation rate and density distribution. A number of works (Kulkarni et al., 2019a; Keating et al., 2020; Nasir & D'Aloisio, 2019) have now found that a later end to reionization at $z \sim 5$ could account for the scatter, as such a scenario would leave large "islands" of neutral hydrogen even at $z < 6$ that could cause significant absorption along quasar sightlines.

Another promising window into reionization comes from the proximity zones of quasars. The high radiation output near to quasars is able to strongly ionise the nearby gas, and thus allow some radiation to escape (Madau & Rees, 2000). This effect is useful for putting constraints on quasar emissivities, and possibly for learning more about the escape fractions of high-redshift galaxies (Bosman et al., 2019).

Lyman-α emitting galaxies

The Lyman-α line is as ubiquitous across astronomy as hydrogen is in the universe. Some early galaxies, like quasars, also emit strongly at the Lyman-α wavelength. These young Lyman-α emitting (LAE) galaxies contain many hot O and B type stars whose radiation is re-processed by hydrogen in the ISM, and emitted through recombinations as Lyman-α radiation (Partridge & Peebles, 1967). To understand how this emission will look when it has escaped from the galaxy's ISM, we need to return to the Voigt profile from Eq. (1.28). Defining the Doppler width parameter $\Delta\nu_D$ as,

$$\Delta\nu_D \equiv \frac{\nu_{21}}{c}\sqrt{\frac{2k_BT}{m_H}}, \tag{1.29}$$

we can then also define a dimensionless frequency parameter,

$$x \equiv \frac{\nu - \nu_{21}}{\Delta\nu_D}. \tag{1.30}$$

The probability of radiation scattering in the ISM decomposes into two distinct regimes (Dijkstra, 2014),

$$\phi(x) \propto \begin{cases} e^{-x^2} & \text{core} : |x| < x_{\text{crit}}, \\ x^{-2} & \text{wings} : |x| \geq x_{\text{crit}} \end{cases} \tag{1.31}$$

[17] A particularly long $\sim$ 110 proper Mpc/h trough, discussed in Becker et al. (2018), was found in quasar ULAS J0148+0600's spectrum.

where $x_{\rm crit}$ is a (temperature-dependent) threshold. Emitted radiation close to line-centre is very likely to be scattered by nearby hydrogen in the ISM, but radiation in the wings is more likely to travel unhindered and escape the galaxy. Each scatter will introduce a small Doppler shift in the radiation frequency due to the relative motion of the scattering hydrogen; as photons random-walk through the ISM this makes it possible for them to reach the wings of the profile and escape. We thus might expect LAEs to have emission profiles with two peaks either side of the Lyman-α frequency, and indeed this is seen both in observed low redshift LAE analogs (Yang et al., 2016) and also numerical radiative transfer simulations (Smith et al., 2018).

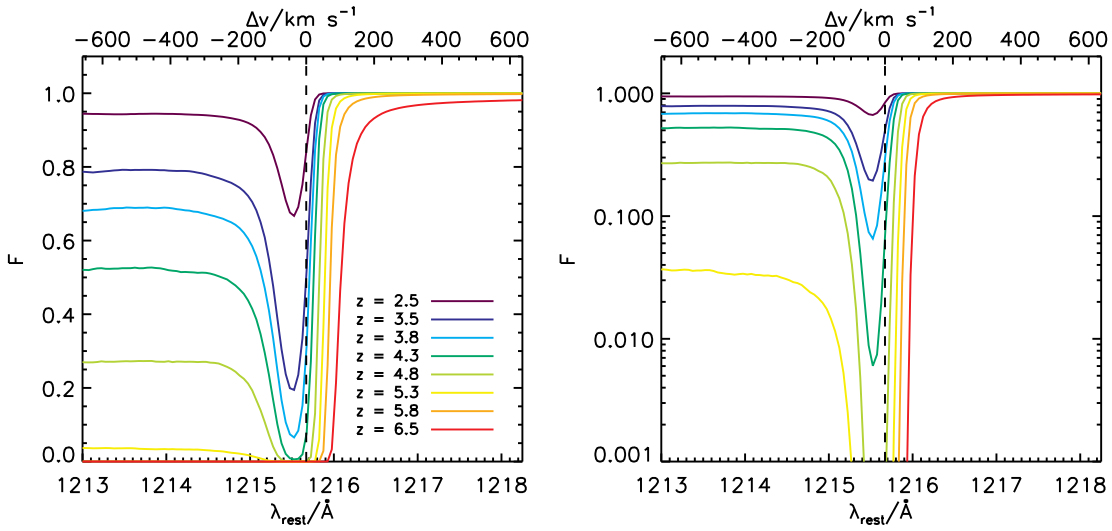


Fig. 1.5 Transmission of radiation at wavelengths around Lyman-α through the IGM, from (Laursen et al., 2011). Different coloured lines indicate different redshifts; the evolution tracks the evolution in the neutral fraction at the end of, and after, reionization. The vertical axes are linear and logarithmic on the left and right panels, respectively.

However, as in the case of the Lyman-α forest, if these galaxies are sufficiently high-redshift then the intervening neutral IGM will be opaque to Lyman-α radiation. In the low-density IGM the attenuation of LAE emission, quantified through the transmission fraction (or *transmissivity*) of photons (Mesinger et al., 2015), is given by,

$$T_{\rm IGM} = \frac{\int {\rm d}\nu \, e^{-\tau(\nu)} \, J(\nu)}{\int {\rm d}\nu \, J(\nu)} \tag{1.32}$$

where $J(\nu)$ is the emission profile of the LAE, and the optical depth is as defined in Eq. (1.26). The transmission, $F(\nu) = e^{-\tau(\nu)}$, of different frequencies through the IGM is shown in Figure 1.5. We note that at high redshifts, before the completion of reionization, the IGM is completely opaque to emission bluewards of Lyman-α.

We can observe LAE galaxies efficiently at high redshifts using narrowband photometric surveys (for example, Hu et al., 1998; Ouchi et al., 2008) which place a filter on the redshifted wavelength and look for emission. The attenuation by the intervening IGM presents another method for constraining the end of reionization: by looking at how the population of observed LAEs evolves with redshift (Dijkstra, 2014), using statistics such as the luminosity and angular correlation functions. Using LAEs for inference is complicated by a number of factors (Dijkstra et al., 2007; Laursen et al., 2011). Exploring some of the difficulties and advantages of using LAEs to understand reionization is the main objective of this thesis, and will be returned to in Chapters 3, 4 & 5.

Recent constraints derived from LAEs include Mason et al. (2018a), who determined that the IGM was still 60% neutral around redshift $z \sim 7$, favouring a later reionization scenario. Furthermore, Becker et al. (2018) used narrowband observations to try to identify LAEs in close proximity to a long absorption trough in the spectrum of quasar ULAS J0148+0600. Keating et al. (2020) used the LAE population model developed in this thesis to show that late reionization scenarios can reproduce the observed "hole" in the LAE surface density around the trough. These observations provide complementary evidence to the CMB and Lyman-α forest results which support a late reionization.

21-cm emission and absorption

Atomic hydrogen also has a hyperfine splitting of the ground state into another electronic (spin-flip) transition with energy 5.87 μeV and corresponding wavelength 21cm, as shown in Figure 1.3. The ratio of the population of the excited state to the ground state can be written in Boltzmann form to define the *spin temperature*,

$$\frac{n_1}{n_0} = \frac{g_1}{g_0} e^{-\Delta E / k_B T_S}, \tag{1.33}$$

where $g_1/g_0 = 3$. The evolution of the spin temperature follows distinct stages in which it is tightly coupled to either the gas or CMB temperatures (Field, 1959).

This transition can also be used as a probe of reionization, and possibly even the end of the earlier dark ages (Fialkov, 2018). As with the Lyman-α line, the 21-cm line is sensitive to the amount of neutral hydrogen; however it is also affected by a complex interaction with the CMB radiation, background Lyman-Werner radiation and gas thermal evolution, resulting in different periods of cosmic history when we might expect to see 21cm emission or absorption compared to the microwave background (see Pritchard & Loeb, 2012, for a review). Unlike for Lyman-α, the IGM is optically thin to 21-cm radiation at all relevant redshifts, potentially providing a window into reionization and beyond. In the optically thin limit the differential brightness temperature compared to the CMB is proportional to (Madau et al., 1997),

$$\delta T_{21} \propto x_{\rm HI} \, (1 + \delta) \left(1 - \frac{T_{\rm CMB}}{T_S} \right). \tag{1.34}$$

Towards the end of reionization it is thought that the spin temperature will have saturated $T_S \gg T_{\rm CMB}$ (Ghara et al., 2015), such that fluctuations in the brightness temperature are driven primarily by the ionisation and density structure of the IGM. There are some heating scenarios which are still physically possible, however, in which the IGM remains cold enough for the spin temperature to not saturate — in this case the fluctuations in the brightness temperature can be enhanced by spin temperature fluctuations where $T_S < T_{\rm CMB}$.

The 21-cm line can be used to probe reionization through a number of different observational methods. One method involves using the sky-averaged "global signal" as a function of frequency (i.e. redshift), a recent example being the approach of the EDGES survey (Bowman et al., 2018). The EDGES team found[18] a large 21-cm absorption signal at a redshift $z \sim 17$, the amplitude and shape of which are difficult to explain with our current understanding of the high-redshift universe. Beyond the global-signal approach it is also possible to explore the full 3D power spectrum of 21-cm emission using radio interferometry, such as with the MWA (Wayth et al., 2018), LOFAR (van Haarlem et al., 2013), HERA (DeBoer et al., 2017) and future SKA (Waterson et al., 2016) interferometers. It may also be possible to map out ionised bubbles using 21-cm tomography (Giri et al., 2018b).

A key difficulty in using the 21-cm line is that the cosmological signal is many orders of magnitude smaller than the foreground signal (McQuinn et al., 2006), for example from synchotron radiation in the Milky Way. There has been considerable work to explore ways to remove (e.g. Morales & Hewitt, 2004; Liu & Tegmark, 2011) or avoid (e.g. Liu et al., 2014) these foregrounds and extract the underlying signal. Another possible way to bypass some of the foreground contamination is to cross-correlate the 21-cm observations with other observables such as LAEs for which we do not expect any correlation with the foreground signal. This is the subject of Chapter 5 of this thesis.

The current picture

In Figure 1.6 we show a collection of some of the latest (at time of writing) observational constraints on reionization. For comparison we also show the theoretical model of Kulkarni et al. (2019a), which has reionization ending at approximately $z \approx 5.3$. As highlighted in the previous sections, recent evidence has been mounting in favour of this "late" reionization scenario. In Figure 1.6 we also show an earlier reionization scenario, the model of Haardt & Madau (2012).

[18] We note that the EDGES result has been questioned by some authors, including Hills et al. (2018), on the grounds of their data analysis in deriving the signal.

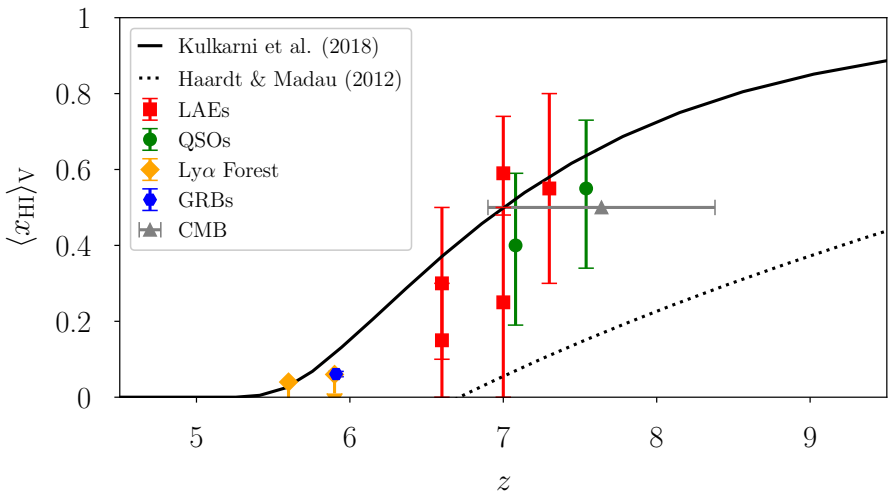


Fig. 1.6 A selection of current constraints on the volume-averaged neutral fraction from measurements of LAEs (red squares, Konno et al., 2018, 2014; Ouchi et al., 2018; Itoh et al., 2018; Mason et al., 2018a); QSO damping wings (green circles, Greig et al., 2017; Bañados et al., 2018); the dark fraction in the Lyman-α forest (yellow diamonds, McGreer et al., 2015); gamma-ray bursts (blue hexagons, Totani et al., 2016); and the CMB (grey triangle, Planck Collaboration, 2018b). The black solid line shows the model of Kulkarni et al. (2019a), which was found to reproduce the opacity fluctuations in the Ly α forest after reionization. For comparison we also show the earlier reionization model of Haardt & Madau (2012) with the black dotted line.

1.3 Thesis Outline

The remaining chapters of this thesis cover research into the theoretical modelling of the EoR, in particular exploring how different reionization histories can imprint an observable signal on the statistics of LAEs and 21-cm emission. It is structured as follows,

- In Chapter 2 we outline the numerical methods underlying the modelling of the IGM and reionization used in later chapters.
- In Chapter 3 we present work exploring the effect of the local environment on Lyα transmission from LAEs.
- In Chapter 4 we present a model for the observed high-redshift LAE population, and compare the predictions of this model with observations.
- In Chapter 5 we present forecasts for observing the LAE-21cm cross-correlation signal as a further means of constraining reionization.
- In Chapter 6 we summarise the work presented here, and reflect on the future outlook of this research field.

Chapter 2

Numerical Methods

2.1 Numerical Simulations

Astrophysics is special amongst the different fields of physics for being concerned with phenomena that occur on scales much beyond the terrestrial. This means the empirical side is driven by observation of the Universe around us, rather than experimentation in the lab. Theoretically, astrophysics is also concerned with large inter-connected systems, often involving very non-linear processes which are difficult to describe analytically. At the intersection of these characteristics lies the solution of numerical computation: experimenting through simulation (Vogelsberger et al., 2020). The basic idea of numerical simulation is to programmatically implement the relevant laws of physics for a given system, such that the system can be evolved from some initial conditions to a later time, and through which a greater understanding of the behaviours of real systems can be developed.

The work of this thesis is concerned with modelling the reionization era using numerical simulations. In particular the later chapters describe the analysis of the Sherwood suite of simulations (implemented and run by Bolton et al., 2017), which were designed to simulate the IGM. The Sherwood simulations were run using the cosmological hydrodynamic code P-GADGET-3 (an evolution of the GADGET code last described in Springel, 2005; Springel et al., 2001). In this chapter we will sketch an outline of the key methods that underpin modern simulations of reionization, including those employed in P-GADGET-3. We will also describe the methods used in later chapters to model reionization with the Sherwood simulations, which form the basis for our analysis of high-redshift LAE galaxy populations and the 21-cm signal.

2.1.1 Approximating physics

The distribution of matter in the Universe is largely governed by gravity and hydrodynamics (Bertschinger, 1998), whilst the ionisation state is controlled by the physics of radiative transfer (Gnedin & Abel, 2001). Simulating each of these aspects of physics requires establishing approxima-

tions to make the problem computationally tractable. We will give an overview of techniques, and highlight their strengths and weaknesses.

An important theme in this chapter will be the question of how to simulate physics that affects a large dynamic range of scales. For example, since the dynamical time of a system scales as $t_{\rm dyn} \propto \rho^{-1/2}$ (Mo et al., 2010), we might want to simulate solar physics with a timestep on the order of minutes, whereas the evolution of a galaxy might be evaluated every 100 million years. The same is true of the physical size of these different systems, which are orders of magnitude apart[1]. It is therefore very challenging to perform a simulation which is able to capture the physics that controls stars and galaxies (and everything in between), and some approximations must be made. Modern cosmological simulations of reionization and galaxy formation are able to resolve down to physical scales of ~ 10 pc at $z = 6$ (Rosdahl et al., 2018; Pallottini et al., 2017). This means that physics below these scales must be further approximated, giving rise to the term *sub-grid physics* (Herring, 1979).

2.1.2 Gravity

As discussed in Section 1.1.1, the physics of gravity is described by General Relativity in terms of the curvature of space-time. However the standard approach to cosmological simulation is to assume a homogeneous expansion and treat gravity in the Newtonian limit, providing a very good approximation[2] (East et al., 2018). In this regime the force on a test mass, m_1, as a result of the gravitational attraction to another test mass, m_2, is given by,

$$\mathbf{F} = m_1 \mathbf{g} = -\frac{G m_1 m_2}{|\mathbf{r}_{12}|^2} \hat{\mathbf{r}}_{12}, \tag{2.1}$$

where $\mathbf{g}$ is the gravitational field strength at m_1's position, r_{12} is the separation between the masses, and $G = 6.67 \times 10^{11}$ m^{3}kg^{-1}s^{-2} is the gravitational constant. The gravitational field strength can also be written in terms of the scalar gravitational potential, Φ, in the form of Poisson's equation,

$$\nabla \cdot \mathbf{g} = \nabla^2 \Phi = -4\pi G \rho, \tag{2.2}$$

where ρ is the density of the attracting body.

To simulate the formation of galaxies we need to solve these gravity equations for the dark matter, gas and stars. In general the differential equations which govern the evolution of physical systems are continuous in space and time. Thus in order to solve for this evolution numerically on a computer, we need to perform some discretisation. Perhaps the simplest approach in the case of gravity is to discretise the density fields into a sample of N particles, such that the force on any particle is given

[1] The solar radius is approximately 7×10^8 m, whilst the radius of the Milky Way is approximately 5×10^{20} m, giving a difference in scale of roughly 12 orders of magnitude.

[2] This neglects the *backreaction* of matter on the evolution of the metric (Macpherson et al., 2019).

by the sum,

$$\mathbf{F}_i = -\sum_{j \neq i}^{N} \frac{G m_i m_j}{|\mathbf{r}_{ij}|^2} \hat{\mathbf{r}}_{ij}. \tag{2.3}$$

This approach, also referred to as *N-body* simulation, is analogous to solving the dynamics of real discrete systems such as the orbital motion of planets. The finer the discretisation (i.e. the smaller the mass of the sampling particles), the more small-scale behaviour can be resolved. We note that the pairwise sum in Eq. (2.3) has a complexity which scales as $\mathcal{O}(N^2)$, which can become computationally prohibitive as the number of particles N increases (Aarseth, 1971). The largest simulations run using this direct pairwise sum are able to use millions of particles (Panamarev et al., 2019).

A number of numerical methods have been developed to optimise this calculation using careful approximations.

- One approach, referred to as the *particle-mesh* (PM) method (Klypin & Shandarin, 1983; Hockney & Eastwood, 1988), involves interpolating particles (which is $\mathcal{O}(N)$ complexity in the particle number) in order to calculate the gravitational potential on a grid, then using this potential grid to displace the particles (which is again $\mathcal{O}(N)$ complexity). Fast Fourier transforms (FFT) can be used to optimise[3] the potential calculation in Eq. 2.2 during the interpolation step. A weakness of this approach is the limit of the grid resolution, below which the force is suppressed. This can be alleviated, however, by modifying the method to include adaptive resolution grids (Couchman, 1991) or direct force summation for short-ranges (such as the P^{3}M code, Efstathiou & Eastwood, 1981; Efstathiou et al., 1985).

- Another set of approaches relies on storing the particle sample in an efficient *tree* data-structure. These hierarchical data-structures provide information about neighbouring particles such that the force calculation can treat nearby groups of particles individually, whilst more distance particles can be treated collectively using their centre of mass, thus reducing the total number of calculations required (Barnes & Hut, 1986). This approximation can be understood in principle by considering the multipole moments of the gravitational potential. At large distances the higher multipoles quickly tend to zero, leaving only the monopole (centre of mass) contribution. This approximation can be improved by including higher order multipoles for the distant particles in the tree (Greengard, 1988). Tree algorithms don't suffer from the same resolution limitations as particle-mesh methods, however they scale as $\mathcal{O}(N \log N)$ in complexity (Appel, 1985).

The P-GADGET-3 code uses a hybrid "TreePM" method which approximates the long-range forces using the particle-mesh approach and the short-range forces using the tree algorithm, thus balancing the limitations of each (Xu, 1995).

[3] The complexity of FFTs scale as $\mathcal{O}(G \log G)$ in the number of grid points, G, (Cooley et al., 1967).

2.1.3 Hydrodynamics

Alongside gravity, for baryonic matter we must also solve for its hydrodynamic evolution. Astrophysical fluids can be approximated as compressible and inviscid (Clarke & Carswell, 2014), such that their dynamics are described by the Euler equations[4],

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} + \rho\nabla\cdot\mathbf{v} = 0, \tag{2.4}$$

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \frac{\nabla P}{\rho} = 0, \tag{2.5}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} + \frac{P}{\rho}\nabla\cdot\mathbf{v} = 0, \tag{2.6}$$

where ρ is the density, $\mathbf{v}$ is the velocity, P is the pressure, E is the total energy, and we note that the above equations are written in Lagrangian form using the convective derivative $\mathrm{d}/\mathrm{d}t = \partial/\partial t + \mathbf{v}\cdot\nabla$. We note that both stars and cold dark matter can be treated as collisionless fluids, such that their evolution is purely determined by gravity as described above.

As in the case of gravity, there are a number of different ways to approach solving these equations numerically (some recent reviews include, Springel, 2010a; Teyssier, 2015). These approaches often include numerical methods for ensuring that shocks – discontinuities in the Euler equations – are treated properly (Schaal et al., 2015). We will give a brief overview of three approaches: the Eulerian *grid-based* methods, the Lagrangian *smoothed-particle hydrodynamics* (SPH), and the quasi-Lagrangian *moving mesh* methods.

Grid-based methods

Historically the first approaches to solving hydrodynamics numerically involved discretising space into a grid, and solving the evolution equations at each point in the grid. This grid tessellation can be *structured* (such as in Ramses, Teyssier, 2002) or *unstructured* (such as in Arepo, Springel, 2010b). A structured grid is composed in a regular pattern that can be indexed by an d-dimensional address, making them efficient to store and access in memory. Unstructured grids on the other hand can form an irregular pattern and thus require their connectivity to be specified, however they can be used to more accurately represent irregular domains. Having specified the grid type, there are then a number of ways to solve differential equations computationally including the *finite difference* and *finite volume* methods. The finite difference methods evaluate derivatives in the evolution equations by differencing the field values at each grid point (Scannapieco & Harlow, 1995); finite volume methods are similar but instead use the field values averaged over the cell volume of each grid point (Moukalled et al., 2016).

[4]The Euler equations can be derived by considering moments of the Boltzmann equation which describes the evolution of the statistical distribution function (Boltzmann, 1872).

One example of a cosmological hydrodynamics code that employs a grid-based approach is the Ramses code (Teyssier, 2002). Ramses uses a regular grid with *adaptive mesh refinement* to better resolve regions of interest. The equations are solved with a conservative finite volume method (Toro, 1997). Domain decomposition is used to divide up the simulated volume and parallelise the calculation across multiple computer processors.

Smoothed-Particle Hydrodynamics

SPH takes a similar approach to N-body gravity methods, discretising the continuous fluid into a sample of interpolating particles. Associated with each particle is a smoothing length, h, and interpolating kernel $W(r, h)$, which can be used to recover the properties of the fluid at any point (Gingold & Monaghan, 1982). For example a generic field, Q, can be recovered by the discrete convolution,

$$Q(\mathbf{r}) = \sum_i Q_i W(|\mathbf{r} - \mathbf{r}_i|, h_i) V_i, \tag{2.7}$$

where the summation is over the particles, and V_i is the volume of the i^{th} particle. The P-GADGET-3 code employs an SPH scheme with cubic spline kernel (Monaghan & Lattanzio, 1985),

$$W(r, h) = \frac{8}{\pi h^3} \begin{cases} 1 - 6\left(\frac{r}{h}\right)^2 + 6\left(\frac{r}{h}\right)^3 & 0 \le \frac{r}{h} \le \frac{1}{2}, \\ 2\left(1 - \frac{r}{h}\right)^3 & \frac{1}{2} < \frac{r}{h} \le 1, \\ 0 & \frac{r}{h} > 1, \end{cases} \tag{2.8}$$

and an adaptive smoothing length which ensures that each kernel volume contains a constant mass for the estimated density. The time integration of the SPH system is commonly performed using symplectic schemes[5] such as the leapfrog method (Quinn et al., 1997).

The resolution of an SPH simulation is a function of the number of particles and the size of the simulated volume. An advantage of the Lagrangian nature of SPH is that denser regions (which may contain e.g. galaxies) are naturally better sampled by more particles. It is also possible to use many of the well-established N-body techniques, such as tree algorithms, to optimise SPH computations. A well known limitation of SPH is its ability to properly capture discontinuities that arise in the Euler equations, such as in astrophysical shocks (Bauer & Springel, 2012). This can be partially mitigated by the introduction of artificial viscosity (Monaghan & Gingold, 1983). Furthermore the discrete sampling introduces an "interpolation" error which produces an inherent noise in SPH results.

Moving mesh methods

A more recent development in this field is the use of moving mesh methods which can be both Eulerian and Lagrangian (Gnedin, 1995). These techniques discretise space into a grid, but allow the grid to move with the fluid. An example code which uses this is the Arepo code (Springel, 2010b)

[5] As many astrophysical processes are not reversible these are not strictly necessary.

which employs an unstructured Voronoi grid using a finite volume solver. The grid-points can be moved, such that Arepo can use either an Eulerian (for a stationary grid) or Lagrangian (for grid points moving with the local velocity) scheme. As in Ramses, a domain decomposition is performed to parallelise the calculation across multiple processors.

2.1.4 Radiative Transfer

So far we have considered methods for simulating the evolution of matter in the Universe, but the other key component that we would like to simulate is radiation. Radiation is governed by the equation of *radiative transfer* (Chandrasekhar, 1960),

$$\frac{\mathrm{d}I_\nu}{\mathrm{d}s} = \frac{1}{c}\frac{\partial I}{\partial t} + \mathbf{n}\cdot\nabla I_\nu = j_\nu - \alpha_\nu I_\nu, \tag{2.9}$$

where I_ν is the specific intensity of radiation at frequency ν, $\mathrm{d}s = c\mathrm{d}t$ is a curvilinear coordinate along a light-ray, $\mathbf{n}$ is the direction of the incoming radiation, j_ν is an emission (source) coefficient, α_ν is an absorption (sink) coefficient, and we have neglected the effect of scattering. We note that the total energy density is given by an integral over solid angle Ω,

$$E_\nu = \int_{4\pi} \mathrm{d}\Omega\, \frac{I_\nu}{c}. \tag{2.10}$$

The specific intensity is a function of frequency, position, direction and time $I_\nu = f(\nu, \mathbf{x}, \mathbf{n}, t)$; this high dimensionality makes radiative transfer a difficult problem to solve. The formal steady-state solution along a single light-ray (again neglecting scattering) is given by (Rybicki & Lightman, 1985),

$$I_\nu(s) = I_\nu(s_0)\, e^{-\tau_\nu(s_0,s)} + \int_{s_0}^{s} \mathrm{d}s'\, j_\nu(s')\, e^{-\tau_\nu(s_0,s')}, \tag{2.11}$$

where the optical depth τ_ν (as seen before in Eqs. 1.24 & 1.26) is defined,

$$\tau_\nu(s', s'') = \int_{s'}^{s''} \mathrm{d}s\, \alpha_\nu(s). \tag{2.12}$$

Solving radiative transfer is key for understanding the photoionisation and photoheating processes at the heart of reionization. Ideally one would solve for the fully coupled evolution of a system; however to make the problem easier we can also make the approximation that the hydrodynamics is decoupled from the radiative transfer. Thus it is possible to run a hydrodynamical simulation first, and evaluate the radiative transfer as a post-processing step (Aubert & Teyssier, 2010).

There are a number of numerical approaches to solving the radiative transfer equation. The *Monte Carlo*[6] approach (for example used in Hyperion, Robitaille, 2011) is similar in philosophy to N-body methods, discretising the radiation field into packets of radiation (Avery & House, 1968). Another method is the *Eddington approximation* (Eddington, 1916), which considers the angular moments of Eq. 2.11,

$$J_\nu = \frac{1}{4\pi} \int_{4\pi} \mathrm{d}\Omega \, I_\nu, \qquad (0^{\text{th}} \text{ moment}) \tag{2.13}$$

$$\mathbf{F}_\nu = \int_{4\pi} \mathrm{d}\Omega \, I_\nu \, \mathbf{n}, \qquad (1^{\text{st}} \text{ moment}) \tag{2.14}$$

$$\mathbb{P}_\nu = \int_{4\pi} \mathrm{d}\Omega \, I_\nu \, \frac{\mathbf{n} \otimes \mathbf{n}}{c}, \qquad (2^{\text{nd}} \text{ moment}) \tag{2.15}$$

where J_ν is the mean specific intensity, $\mathbf{F}_\nu$ is the radiation flux, and $\mathbb{P}_\nu$ is the radiation pressure (sometimes defined in terms of the *Eddington tensor*, $\mathbb{P}_\nu = \mathbb{D}_\nu E_\nu$). Energy and radiative flux conservation couple these three moments together, but the radiation pressure tensor needs to be specified to close the equations (González et al., 2007).

As an example, the Aton code (Aubert & Teyssier, 2008) uses the moments-based approach with an optically thick approximation called the "M1 closure" (Levermore, 1984) to specify the second moment in the reionization regime. Alternative closures are accurate in other regimes, such as the optically thin closure of Gnedin & Abel (2001). Having specified the M1 closure, Aton simulates the evolution of radiation in three stages: (i) the radiation field is generated from the source population of a given simulation, (ii) the moments-based radiative transfer equation is solved using a conservative finite-volume scheme, and (iii) finally the thermochemical state of the simulation is solved, determining the effects of photoionisation and photoheating.

Calculating full radiative transfer as described above can be computationally prohibitive. A number of "semi-numerical" techniques have been created which make further simplifying assumptions in order to calculate the effect of radiative transfer more efficiently. We will return to a particular *excursion set*-based approach in Section 2.2.2.

We note that in the context of observations of LAEs, scattering of Lyman-α radiation by the IGM cannot be neglected as in Eq. 2.11. In Section 3.3 we discuss techniques for modelling this behaviour numerically.

[6]The first reported use of the Monte Carlo method was to simulate nuclear fission processes, originally an idea of Stanislaw Ulam (Metropolis, 1985). These computer simulations were run by John von Neumann on the ENIAC machine (Eckhardt, 1987), one of the first general-purpose electric computers.

2.1.5 Sub-grid Physics

In reality the physical processes described above are coupled to each other across large spatial and temporal scales, as well as to other important forces such as magnetism and processes such as star formation which we have not described. In order to solve for the evolution of astrophysical systems, these processes are approximated and in some cases decoupled in order to ensure the problem is computationally tractable.

One important feature of all of the methods described above is the limitation of resolution and thus the physics which is missed below this limiting scale. Current simulations of galaxy formation and the evolution of the IGM need to capture spatial scales much larger than those of, for example, individual star formation. In order to include these processes, many simulations employ "sub-grid" models which describe the average behaviour at the resolution scale. A common approach to modelling star formation is to adapt the star formation scaling of Schmidt (1959); Kennicutt (1998), which relates star formation to surface brightness, and thus a gas density threshold (Springel & Hernquist, 2003). Chemical evolution and heating processes can be similarly modelled (Ferland et al., 2017; Ferland, 1983).

As with all numerical approaches, the balance of approximations needs to be chosen to best fit the desired problem. In the simulations of the IGM described in later chapters, the QUICK_LYA star formation prescription (Viel et al., 2004) is used in P-GADGET-3. This is a sub-grid model which converts cold gas with temperatures lower than $T < 10^5$ K and overdensities higher than $\Delta > 10^3$ into star particles. Although a simplistic model for star formation, which does not accurately capture feedback processes, this prescription allows higher resolution simulations to be run in order to better capture the low-density IGM. Of particular importance in simulations of the IGM during reionization is the ability to capture overdense pockets of gas with high recombination rates which are able to "self-shield" against the background ionising radiation, thus remaining neutral (Miralda-Escudé et al., 2000; Rahmati et al., 2013; Chardin et al., 2018). We note that the the mean-free path of ionising photons through the IGM is sensitive to this self-shielded gas, which occurs at overdensities $\Delta \sim 10$–100 (Pawlik et al., 2010), and is therefore not affected by the QUICK_LYA removal of $\Delta > 10^3$ gas.

2.1.6 Historical Aside

The above sections gave an overview of some of the key physical processes that we can attempt to simulate numerically in order to model astrophysical phenomena, such as the formation of galaxies. As discussed, many algorithmic advances have been made to optimise these calculations. It is also interesting, however, to consider the wider historical context of computing and how advances in technology have impacted the field of computational astrophysics. We will briefly discuss the

development of processor speeds[7] and memory capacity, and use the size of N-body calculations as a rough measure of capability (see Figure 2.1).

One of the earliest astrophysical simulations actually predates the first general purpose computers: Holmberg (1941) used an elaborate setup of $N = 74$ lightbulbs and photocells to simulate the merger of two nebulae[8]. The development of general purpose computers was accelerated during the Second World War, with the invention of machines such as ENIAC, EDVAC, the Manchester Baby and EDSAC (Wilkes, 1956) in the late 1940s that used vacuum tubes for logic circuitry. The memory capacity of these early computers was on the order of kilobytes, and the time to perform a single multiplication was a few seconds. The majority of simulations run with these machines were geared towards ballistics calculations. A second generation of computers developed around transistors in the late 1950s, including mainframes such as the IBM 7090, the DEC PDP-1 and IBM's System/360 line. These machines used *core* memory on the order of megabytes and were able to process calculations at a rate of ~ 100 kiloflops. Academic use of computers had grown significantly by this point, with some astrophysics departments investing in their own machines (Hoyle, 1994). For example, (Aarseth, 1963, 1966) used the IBM 7090 and 7094 computers to run N-body simulations of galaxy clusters with $N = 25–100$ particles. The development of computers accelerated over the following decades with the invention of integrated circuits (Kilby, 1976), and the introduction of smaller computers such as the minicomputers of the 1960s, the microcomputers of the 1970s, and finally the personal computers of the 1980s. The cutting edge astrophysical simulations of the time were run on supercomputers, such as the Cray X-MP, which introduced multiprocessor and vector ("single instruction, multiple data") architectures that allowed calculations to be performed in parallel (Efstathiou, 1986). These supercomputers were able to compute at 10–100 megaflops, and were able to run N-body simulations with $N = 1000–100,000$ particles. For example, James & Sellwood (1978) ran a simulation of galactic discs using 25,000 particles on the CDC 7600 (boasting a peak performance of 36 megaflops). Almost three decades later in the mid 2000s, the Millennium Run simulation (Springel et al., 2005) was run on 512 processors of the IBM pSeries 690 parallel computer using just over a billion particles, requiring a total (wall clock) time of 28 days. It is worth noting that individual processor speeds have increased exponentially (as a result of the doubling roughly every 2 years of the number of transistors that can be built into an integrated circuit, as observed by Moore, 1998). Alongside this, greater performance has been achieved by increased parallelisation (at the instruction level, processor level, and to some extent at the operating system level).

[7] A common measure of processing speed is the number of floating point operations per seconds ("*flops*").

[8] The lightbulbs represent a discrete sampling of the mass field, and their intensity (which, like Newtonian gravity, falls off as $\propto r^{-2}$) was used as an analog for the gravitational field.

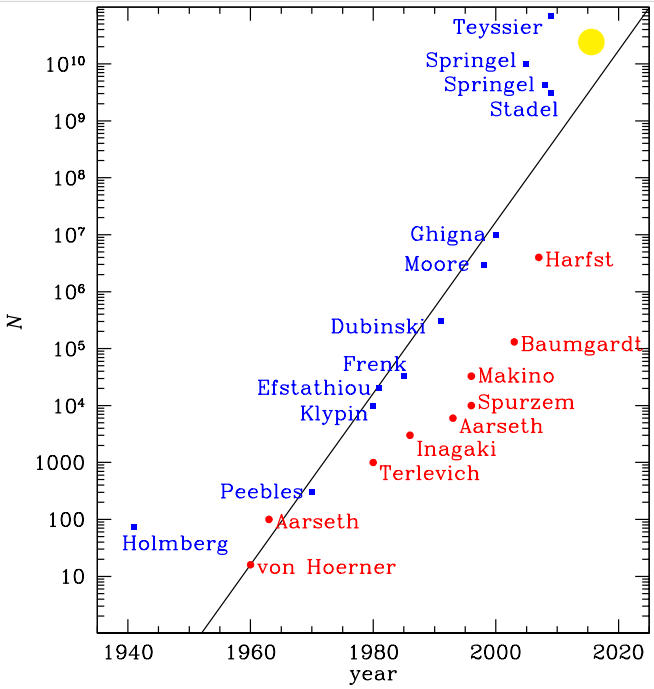


Fig. 2.1 The Moore's law of N-body simulation, from Dehnen & Read (2011). Blue and red points indicate collisionless (dark matter, stars, etc.) and collisional simulations. For comparison, the Sherwood simulations employed in later chapters were run by Bolton et al. (2017) with $N = 2 \times 2048^3$ particles, indicated here by the yellow circle.

A recent example of a high resolution cosmological simulation is the Outer Rim project (Heitmann et al., 2019), which was run on the Mira supercomputer with over a trillion particles. Mira is built using IBM's BlueGene/Q architecture, achieving petaflop performance. The goal of exascale high performance computing is set to be reached within the decade. Modern and future astrophysical simulations need to take advantage of parallel and heterogeneous architectures, for example employing graphical processing units (GPUs) for specific compute heavy tasks.

2.2 Reionization History Modelling

In the preceding section we gave a general overview of the numerical ingredients involved in simulating the IGM. We will now discuss how we model the progress of reionization using the Sherwood suite of simulations. As outlined in Section 1.2, the evolution of reionization is determined by the distribution of both the neutral hydrogen in the IGM, and the sources which radiate ionising photons. In order to implement reionization modelling in the Sherwood simulations we need to make some assumptions:

- The radiative transfer of ionising photons is decoupled from the hydrodynamic evolution of the IGM gas. This means we are neglecting the hydrodynamic response of the gas to the radiation field, but allows us to *post-process* snapshots of the hydrodynamic simulation with a range of

different reionization scenarios. This assumption is valid in the regime of reionization, where thermal pressure dominates over the hydrodynamic response due to passing ionisation fronts (Bolton et al., 2017; Finlator et al., 2012).

- A reionization history, in the form of the evolution of the global average ionised fraction $\langle x_{\mathrm{H\,II}}(z)\rangle$. As there is still some uncertainty in the exact nature of the sources of reionization, we choose to first assume a particular reionization history and then determine the required source population that can produce such a scenario. This allows us to explore reionization histories that are consistent with current observational constraints.

- Given a reionization history, in order to determine the ionisation structure of the IGM we need to assume an appropriate source distribution. The Sherwood simulations do not resolve the formation of galaxies, instead employing the QUICK_LYA star formation prescription for computational efficiency. We therefore use the dark matter halo population of the simulation, and assume an ionising emissivity per halo mass. The choice of this emissivity function determines how clustered ionising sources are, and hence can be tweaked to simulate either a galaxy or quasar driven reionization (see for example Kulkarni et al., 2017).

In the work described in later chapters, we employ a semi-numerical "excursion-set" scheme to determine the ionisation state of the IGM. This scheme is computationally efficient, and has been found to agree well with full radiative transfer simulations (c.f. those described in Section 2.1.4). As a comparison in Chapter 4 we also use the Aton radiative transfer code to check that our results are not sensitive to this approximate scheme.

2.2.1 The Sherwood Suite

The Sherwood simulation suite (see Bolton et al., 2017, for full details) includes a range of cosmological hydrodynamic simulations of different resolutions. Each simulation was run with a modified version of P-GADGET-3 as a cubic, periodic box (with side length L_{sim}) containing N_{sim} particles of dark and visible matter combined. Throughout this thesis we will identify each simulation by these two characteristics; of principle interest to later chapters are the $L_{\mathrm{sim}} = 160$ cMpc/h and $L_{\mathrm{sim}} = 320$ cMpc/h boxes, each with $N_{\mathrm{sim}} = 2 \times 2048^3$ particles. The cosmological parameters of the simulation were chosen to match the ΛCDM *Planck+WP+highL+BAO* parameters (Planck Collaboration et al., 2014): $h = 0.678$, $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_b = 0.0482$, $\sigma_8 = 0.829$, $n = 0.961$, and $Y_{\mathrm{He}} = 0.24$. Dark matter haloes were identified on-the-fly using a Friends of Friends (FOF) algorithm.

The gas temperature evolution is modelled assuming a spatially uniform UV background as suggested by Haardt & Madau (2012)[9]. This treats reionization as occurring instantaneously at $z_{\mathrm{re}} = 15$, which has some impact on the pressure-smoothing of the gas at lower redshifts. The

[9] A modified evolution is used at low redshifts (not considered in this thesis) to better match the temperature measurements of Becker et al. (2011).

temperature of the IGM is calculated assuming the gas is optically thin and in ionisation equilibrium with the uniform background. We note that when employing full radiative transfer post-processing with Aton (see Section 4.6.1) this temperature distribution is recalculated and properly coupled to the radiation field.

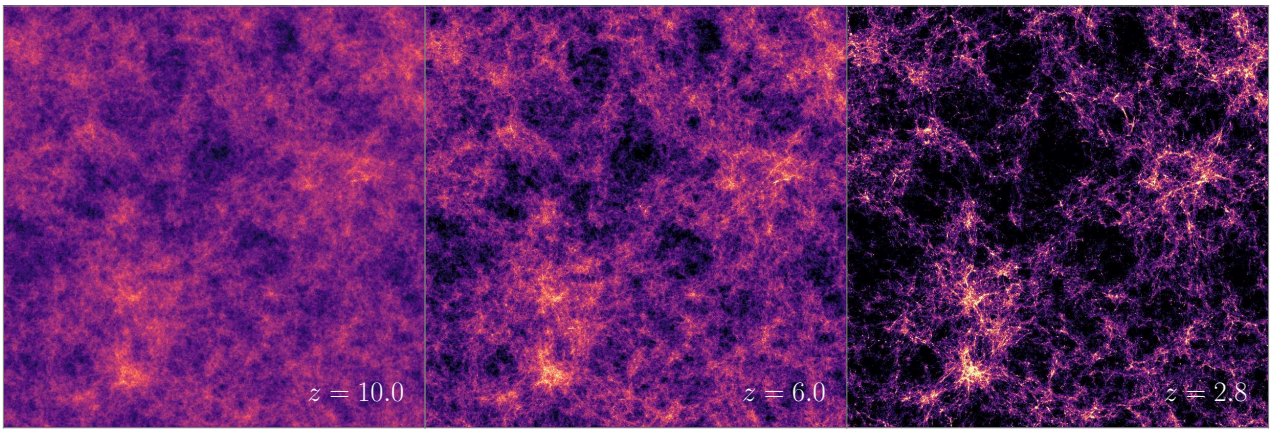


Fig. 2.2 Projections of the density field of the Sherwood $L = 160$ cMpc/h, $N = 2 \times 2048^3$ simulation, from left to right showing redshifts $z = 10, 6$ and 2.8.

We note in passing that the Sherwood simulation suite has been used for a range of science projects. These include modelling the Lyman-α forest after reionization at $z \leq 5$ (Bolton et al., 2017; Kulkarni et al., 2019a; Keating et al., 2020; Gaikwad et al., 2020), the 21-cm signal during reionization (Kulkarni et al., 2016, 2017; Dumitru et al., 2019), and high-redshift Lyman-α intensity mapping (Witstok et al., 2019). In Figure 2.2 we show projections of the density field from the $L = 160$ cMpc/h, $N = 2 \times 2048^3$ simulation. The density field is derived from the particle field using the interpolating kernel (as described in Eq. 2.7).

2.2.2 The Excursion Set Method

Solving for the ionisation state of the IGM during reionization can be achieved through a photon counting scheme, which for historical reasons we will refer to as an *excursion set* approach, pioneered by Furlanetto et al. (2004c,d) and further developed in Mesinger & Furlanetto (2007); Choudhury et al. (2009); Mesinger et al. (2011); Santos et al. (2010); Hassan et al. (2016). Instead of solving the full radiative transfer of ionising radiation, this method reduces the problem to a photon and hydrogen number balancing exercise.

In order to analytically calculate the distribution of ionised H II regions in the IGM, Furlanetto et al. (2004c) used an approach analogous to the Press & Schechter (1974) calculation of the halo mass function:

- In the halo formation case (Press & Schechter, 1974), we consider the density field smoothed on a particular mass scale, M, and find collapsed regions where the density field fluctuates above a critical threshold, δ_c.
- In the ionised bubble formation case (Furlanetto et al., 2004c), we are instead interested in whether a given region is able to 'self-ionise'; in other words whether on some radial scale R, there are enough luminous sources within a region to ionise all of its hydrogen atoms.

To evaluate this excursion condition for ionised regions, we can take a halo-model based approach and assume that ionising emissivity scales with halo mass (Choudhury et al., 2009),

$$N_\gamma = c_\gamma M_h, \tag{2.16}$$

where c_γ is an efficiency parameter. In the case of an analytic calculation of the distribution of ionised regions (as in Furlanetto et al., 2004c), one might combine Eq. (2.16) with a halo mass function such as derived using the extended Press-Schechter formalism (Bond et al., 1991). However in our case we are trying to determine the location of ionised regions within our hydrodynamic simulations, hence we use the halo catalogue of the simulation.

Given this halo-based model for the ionising emissivity, and the hydrogen density distribution of the simulation, we can evaluate the excursion condition to determine the location of ionised regions. A spherical region of radius R centred on the $\mathbf{i}^{\text{th}}$ cell can 'self-ionise' if (Choudhury et al., 2009),

$$\langle n_\gamma(\mathbf{i})\rangle_R > \langle n_{\text{H}}(\mathbf{i})\rangle_R \, (1 + \overline{N}_{\text{rec}}), \tag{2.17}$$

where $\langle\,\rangle_R$ indicates a spatial average within a spherical region of radius R centred on the cell. The left hand side of this equation corresponds to the average (comoving) number density of ionising photons in the vicinity of the cell, whilst the right hand side gives the average number density of hydrogen atoms. The factor $(1 + \overline{N}_{\text{rec}})$ accounts for an average of $\overline{N}_{\text{rec}}$ recombinations per hydrogen atom. Determining the ionisation state then reduces to testing this condition for every cell, scanning across relevant radial scales to count the average number of photons. In analogy to the halo formalism of Bond et al. (1991), we are interested in the largest scale at which this condition holds true in order to properly account for ionising photons from neighbouring regions (Furlanetto et al., 2004c).

Although we can derive n_{H} directly from the simulation, we still need to determine c_γ and $\overline{N}_{\text{rec}}$. Rather than modelling these parameters from first principles, we instead calibrate them to give a desired reionization scenario (specified by the average neutral fraction $\langle x_{\text{HI}}\rangle$). Rewriting Eq. 2.17 in terms of the fraction of mass contained in halos f_h, we have an ionisation condition of (Choudhury et al., 2009),

$$\zeta_{\text{eff}}\, f_h(R) \geq 1, \tag{2.18}$$

where $\zeta_{\rm eff} = c_\gamma m_H \left((1+\overline{N}_{\rm rec})(1-Y_{\rm He})\right)^{-1}$ is an efficiency parameter, and the fraction of mass within haloes is given by,

$$f_h(R) = \frac{1}{\overline{\rho}(R)} \int_{M_{\rm min}}^{\infty} dM \left.\frac{dn}{dM}\right|_R M, \tag{2.19}$$

where $M_{\rm min}$ is the minimum mass halo expected to form galaxies and produce ionising radiation (Choudhury et al., 2008). The mean density $\overline{\rho}(R)$ is calculated within a spherical region of radius R. Specified in this way, constructing the large-scale ionisation structure of the IGM during reionization involves two stages: (i) determining the collapsed mass fraction from the halo catalogue, (ii) calibrating the single parameter $\zeta_{\rm eff}$ to give a desired neutral fraction (see Section 2.2.4).

The use of this semi-numerical excursion set approach for modelling reionization observables, such as the 21-cm signal, has been compared to full radiative transfer calculations (see for example Zahn et al., 2011; Majumdar et al., 2014). These comparisons found reasonable agreement between the methods when calculating observables such as the 21-cm power spectrum, at scales relevant to reionization. A key advantage of the semi-numerical approaches is their greatly reduced computational cost. We note however that the excursion set method as presented above does not properly conserve photon number (Choudhury & Paranjape, 2018); improved schemes have been suggested to account for this.

The linear scaling of emissivity with halo mass in Eq. (2.16) is motivated by earlier work such as Trenti et al. (2010) that found a linear scaling is able to reproduce observed high-redshift galaxy UV luminosity functions. Furthermore, Chardin et al. (2015) employed a linear scaling with full radiative transfer simulations of the end of reionization and was similarly able to reproduce observations. We note however that recent high redshift galaxy observations (such as Mason et al., 2015) may indicate a need for emissivities with a non-linear mass scaling. The presence of galactic outflows and strong feedback may also alter the mass dependence (Finlator et al., 2011). Finally we note that Kulkarni et al. (2016) also employed the reionization framework considered here; they explicitly compared linear and non-linear models, finding minimal impact on their results.

2.2.3 Photoionisation equilibrium

The excursion set modelling described above delineates ionised regions within the IGM. However it does not accurately model the partial ionisation state of gas within ionised regions. We therefore recalculate the ionisation state of gas flagged as ionised by the excursion set method, by assuming it is in photoionisation equilibrium with a uniform UV background (Meiksin, 2009),

$$\frac{{\rm d}x_{\rm HI}}{{\rm d}t} = -x_{\rm HI}\,\Gamma_{\rm HI} + x_{\rm HII}\,n_e\,\alpha_{\rm R}(T), \tag{2.20}$$

where n_e is the free electron number density, $\alpha_{\rm R}(T)$ is the recombination rate and $\Gamma_{\rm HI}$ is the background photoionisation rate. Since we are interested in the ionisation state of highly ionised regions, in

equilibrium this rate equation reduces to,

$$x_{\rm HI}^{\rm eq} = \frac{n_e\ \alpha_{\rm R}(T)}{\Gamma_{\rm HI} + n_e\ \alpha_{\rm R}(T)}, \tag{2.21}$$

The assumption of a uniform UV background is valid towards the end of reionization as ionised bubbles begin to percolate, however it may break down at high redshifts near the start of reionization. Furthermore it neglects the variation in the ionisation field close to sources (see Section 3.6.2) and dense clumps of gas (see Section 2.2.5).

We note that the background photoionisation rate is not a free parameter, but instead depends on the ionising emissivity and mean free path (Kuhlen & Faucher-Giguère, 2012; Becker & Bolton, 2013),

$$\Gamma_{\rm HI} = \frac{(1+z)^2 \sigma_{\rm H} \lambda_{\rm mfp}}{\langle x_{\rm HII} \rangle_{\rm V}} \left(\frac{\alpha_s}{\alpha_b + 3} \right) \langle \dot{n}_{\rm ion} \rangle, \tag{2.22}$$

where $\sigma_{\rm H}$ is the hydrogen photoionisation cross-section at 912 Å, $\lambda_{\rm mfp}$ is the mean free path of ionising radiation at the same wavelength, and the bracketed factor includes the spectral indices for ionizing sources α_s and the ionizing background α_b[10]. The mean free path itself depends on the distribution of ionised regions $x_{\rm HII}(\mathbf{i}, z)$ and is particularly sensitive to small-scale neutral clumps, such as gas that is able to self-shield and remain neutral. Throughout this work we use the ionising spectrum of Haardt & Madau (2012) for the bracketed term on the right hand side of Eq. (2.22).

2.2.4 Calibration to a specified reionization history

The above framework can be used to generate an accurate ionisation field for the IGM during reionization, dependent on the desired reionization history which is specified by the average neutral fraction evolution, $\langle x_{\rm HI}(z) \rangle$ (or indeed the ionisation fraction $x_{\rm HII} = 1 - x_{\rm HI}$). This single input is used to determine the required $\zeta_{\rm eff}$ and $\Gamma_{\rm HI}$ for a self-consistent ionisation state.

When applying this framework to model reionization in later chapters, we use the *mass-averaged* ionisation fraction, which can be derived from the gridded simulation,

$$\langle x_{\rm HII} \rangle_M = \frac{1}{M_{\rm tot}} \sum_{\mathbf{i}} M(\mathbf{i})\ x_{\rm HII}(\mathbf{i}) = \frac{1}{\rho_{\rm tot}} \sum_{\mathbf{i}} \rho(\mathbf{i})\ x_{\rm HII}(\mathbf{i}), \tag{2.23}$$

where the sum is over all $N_{\rm grid}$ cells of the uniform simulation grid. We note the common alternative average is *volume-weighted* (also called the volume-filling factor),

$$\langle x_{\rm HII} \rangle_V = \frac{1}{N_{\rm grid}} \sum_{\mathbf{i}} x_{\rm HII}(\mathbf{i}), \tag{2.24}$$

[10]Not to be confused with the case-B recombination coefficient α_B.

In order to derive the required $\zeta_{\rm eff}$ for a particular redshift snapshot of our simulation, we start with a guessed value and determine the resulting mass-averaged ionisation fraction using Eq. 2.23. We then iteratively modify $\zeta_{\rm eff}$ until the desired $\langle x_{\rm HII} \rangle_M$ is found.

To determine $\Gamma_{\rm HI}$ we iteratively solve the reionization equation[11] for our desired reionization history (Kulkarni et al., 2016; Choudhury et al., 2015; Choudhury, 2009),

$$\frac{{\rm d}\langle x_{\rm HII} \rangle_M}{{\rm d}t} = \frac{\langle \dot{n}_{\rm ion} \rangle - \langle \dot{n}_{\rm rec} \rangle}{\langle n_{\rm H} \rangle} \tag{2.25}$$

As with determining $\zeta_{\rm eff}$, this procedure begins with a guessed value for the ionising emissivity $\langle \dot{n}_{\rm ion} \rangle$, which combined with the mean free path $\lambda_{\rm mfp}$ can be used to derive the photoionisation rate via Eq. (2.22). The derived photoionisation rate $\Gamma_{\rm HI}$ can be used to determine the partial ionisation state of the gas within ionised regions using Eq. (2.21). Similarly we can estimate the comoving recombination rate,

$$\langle \dot{n}_{\rm rec} \rangle = \frac{1}{N} \sum_{\rm i} \alpha_R (1+z)^3 \, n_{\rm e}({\rm i}) \, n_{\rm HII}({\rm i}) \tag{2.26}$$

$$\simeq \frac{1}{N} \sum_{\rm i} \alpha_R (1+z)^3 \, f_{\rm e} \, n^2_{\rm HII}({\rm i}), \tag{2.27}$$

where f_e is the number of electrons per hydrogen nucleus. Hence we can determine both sides of Eq. (2.25) using our input reionization history and the simulation (with a given value of $\langle \dot{n}_{\rm ion} \rangle$). We then iteratively update the value of $\langle \dot{n}_{\rm ion} \rangle$ to balance Eq. (2.25), thus calibrating a value of $\Gamma_{\rm HI}$ consistent with our desired value of $\langle x_{\rm HI} \rangle$.

We note that the mean free path in Eq. (2.22) can be estimated from the simulation. Using a sample of sightlines through the simulation volume, the average transmission can be determined and used to fit the relation (Kulkarni et al., 2016; Rybicki & Lightman, 1985),

$$\langle \exp\left(-\tau_{912}\right) \rangle = F_0 \exp\left(-\frac{x}{\lambda_{\rm mfp}}\right), \tag{2.28}$$

where x is the position along the sightline. Throughout this work we instead choose to employ theoretical models for the mean free path evolution derived from high resolution radiative transfer simulations. This choice was found to be more stable when numerically solving Eq. (2.24) for $\Gamma_{\rm HI}$ in the work described in Chapter 3. To test that this was not sensitive to the resolution needed to properly capture self-shielded regions, we calculated the mean free path from the simulation using our calibrated photoionisation rates. These calculations were indeed converged with respect to the predicted values from the models. This suggests that although we have to fix the mean free path for the calibration, we do properly resolve the self-shielded systems.

[11] This is the mass-averaged equivalent of Eq. 1.22.

This framework allows us to choose a given reionization scenario, specified by the average ionisation evolution, and impose this history onto the Sherwood simulations. In later chapters we use this to compare different histories which bracket current constraints, in order to explore how different observables such as LAEs and the 21-cm signal are impacted by reionization.

2.2.5 Self-Shielding

The assumption of a uniform background photoionisation rate in the above framework has an important limitation. Dense clumps of gas that are able to self-shield against the background ionising radiation will not be properly captured when solving for photoionisation equilibrium with Eq. (2.21). These dense neutral clumps can be classified as Lyman-limit systems (LLS) with column densities between $17.2 < \log_{10}(N_{\rm HI}/{\rm cm}^{-2}) \leq 20.3$ (Fumagalli et al., 2016) and damped Lyman-α absorbers (DLA) with column densities $\log_{10}(N_{\rm HI}/{\rm cm}^{-2}) > 20.3$ (Pontzen et al., 2008). Self-shielding can occur when these pockets of dense gas become optically thick to the background radiation; we note that systems need column densities of $N_{\rm HI} \sim 10^{18}$ cm^{-2} in order to reach optical depths of unity (Erkal, 2013).

Given this dependence on the gas column density, it may appear that full radiative transfer is needed to determine whether self-shielding might occur in a our simulated distribution of gas. However, straightforward physical arguments were suggested by Schaye (2001) which allow the onset of self-shielding to be instead related to the local gas density. Schaye (2001) considered density perturbations in local hydrostatic equilibrium (which correspond to absorbing gas clouds), and argued that for sightlines passing through these perturbations the characteristic size of an absorber will be of order the Jeans length (c.f. Eq. 1.13). Furthermore, along these sightlines a characteristic density exists which corresponds to the column density weighted density of the absorber. For a wide-range of absorber density profiles the maximum density dominates the column density. Hence a typical column density for an absorber can be derived from the product of the local Jeans length and the local density. Inverting these arguments, we can predict the typical density at which self-shielding begins (Rahmati et al., 2013; Furlanetto et al., 2005)

$$\Delta_{\rm ss} = 36 \left(\frac{\Gamma_{\rm HI}}{10^{-12}{\rm s}^{-1}}\right)^{2/3} \left(\frac{T}{10^4{\rm K}}\right)^{2/15} \left(\frac{\mu}{0.61}\right)^{1/3} \times \left(\frac{f_{\rm e}}{1.08}\right)^{-2/3} \left(\frac{1+z}{8}\right)^{-3}, \tag{2.29}$$

where μ is the mean molecular weight. It is therefore possible to approximate the effect of self-shielding using prescriptions based on the local gas density.

In our simulations we account for this self-shielded gas using a sub-grid prescription derived from the high-resolution radiative transfer simulations of Chardin et al. (2018). This prescription represents a redshift-dependent version of the prescription presented by Rahmati et al. (2013). Depending on the gas overdensity, the local photoionisation rate is altered compared to the uniform background

according to,

$$\frac{\Gamma_{\mathrm{HI}}}{\Gamma_{\mathrm{HI,\,global}}} = (1 - f(z)) \left[1 + \left(\frac{\Delta_{\mathrm{H}}}{\Delta_{\mathrm{ss}}}\right)^{\beta(z)}\right]^{\alpha_1(z)} + f(z) \left[1 + \left(\frac{\Delta_{\mathrm{H}}}{\Delta_{\mathrm{ss}}}\right)\right]^{\alpha_2(z)}, \tag{2.30}$$

where Δ_{ss} is the overdensity threshold for self-shielding, and f, β, α_1, α_2 are the redshift dependent parameters found by Chardin et al. (2018). The self-shielding threshold found by Chardin et al. (2018) is in reasonable agreement with the parametrisation based on the local Jeans length in Eq. (2.29) (Schaye, 2001; Furlanetto et al., 2005).

With this prescription, the modified local photoionisation rate can be used with Eq. (2.21) to determine the ionisation state of self-shielded gas within ionised regions.

Chapter 3

Lyman-α Emitters Gone Missing: the Different Evolution of the Bright and Faint Populations

In this chapter we model the transmission of the Lyman-α line through the circum and intergalactic media around dark matter haloes expected to host Lyman-α emitters (LAEs) at $z \geq 5.7$, using the high-dynamic-range Sherwood simulations. We find very different CGM environments around more massive haloes ($\sim 10^{11}\mathrm{M}_{\odot}$) compared to less massive haloes ($\sim 10^{9}\mathrm{M}_{\odot}$) at these redshifts, which can contribute to a different evolution of the Lyα transmission from LAEs within these haloes. Additionally we confirm that part of the differential evolution could result from bright LAEs being more likely to reside in larger ionised regions. We conclude that a combination of the CGM environment and the IGM ionisation structure is likely to be responsible for the differential evolution of the bright and faint ends of the LAE luminosity function at $z \geq 6$. More generally, we confirm the suggestion that the self-shielded neutral gas in the outskirts of the host halo can strongly attenuate the Lyα emission from high redshift galaxies. We find that this has a stronger effect on the more massive haloes hosting brighter LAEs. The faint-end of the LAE luminosity function is thus a more reliable probe of the average ionisation state of the IGM. Comparing our model for LAEs with a range of observational data we find that our results favour 'late' reionization histories, in which reionization finishes rather rapidly at around $z \simeq 6$.

Declaration

The work presented in this chapter was published as a paper titled "*Lyman-α emitters gone missing: the different evolution of the bright and faint populations*", published in Monthly Notices of the Royal Astronomical Society, Volume 479, Issue 2, p.2564-2587. This work was done in collaboration with Girish Kulkarni, Martin G. Haehnelt, Tirthankar R. Choudhury and Ewald Puchwein. The underlying Sherwood simulations that form the basis for the modelling in this work were performed by a prior collaboration (Bolton et al., 2017). Otherwise the modelling and analysis were entirely performed by this author, with guidance provided by the collaborators.

3.1 Background

A notable observation made in recent years is the dramatic decline in the space density of Lyα emitting galaxies (LAEs) beyond $z > 6$ (Kashikawa et al., 2006; Ouchi et al., 2010; Hu et al., 2010; Konno et al., 2014), compared to continuum selected galaxies (Bouwens et al., 2015; Stark et al., 2011; Pentericci et al., 2014; Schenker et al., 2012). Note that at lower redshifts ($3 \lesssim z \lesssim 5$, after hydrogen reionization), however, the LAE luminosity function shows little evolution (Hu et al., 1998; Ouchi et al., 2008). With an increasingly neutral fraction of hydrogen beyond $z \sim 6$, we expect more of the Lyα emission to be absorbed and scattered by the IGM, and hence a reduction in observed flux compared to the continuum. This has been used to obtain model-dependent constraints on the evolution of the neutral hydrogen fraction. For example Ota et al. (2017) used the model of Santos (2004) to convert a Lyα transmission ratio into a fraction $x_{\rm HI}^{z=7} \gtrsim 0.3 - 0.4$.

There have been a number of analytic and numerical models developed to explain the apparent rapid decline of Lyα emission from galaxies; for example taking into account the role of dust and reionization (Dayal et al., 2009), of self-shielded absorbers (Bolton & Haehnelt, 2013; Choudhury et al., 2015), the infall of the CGM onto the host haloes (Sadoun et al., 2017), or ruling out the role of IGM attenuation as a sole factor (Mesinger et al., 2015).

One of the difficulties in explaining this decline is the dependence of the IGM transmission on the Lyα emission line profile of the galaxy, which is complicated by the Lyα radiative transfer out of the galaxy's interstellar medium (ISM) (as discussed in Section 1.2.2). It has been found empirically that the peak of the emission profile is often offset redwards from the Lyα frequency (Erb et al., 2010). Studies at lower redshifts have found correlations between this offset and emission properties such as line magnitude or equivalent width (Yang et al., 2016). For high redshifts the usual reference lines for determining this offset (such as [O III] or Hα) are not observable with ground based telescopes. This leaves either using scaling relationships from low redshift observations (Erb et al., 2014) or, if available, using detections of lines such as C III]λ1909 (Stark et al., 2015).

Theoretical modelling of the Lyα emission profile is made difficult by the resonant nature of the line, resulting in emission profiles that are strongly affected by the ISM (Zheng et al., 2010). Use of Monte Carlo radiative transfer codes (Orsi et al., 2012; Gronke & Dijkstra, 2016) and analytic methods (Dijkstra et al., 2006) has led to simple parametrised models of the emission profile such as the shell model (Gronke et al., 2015), but see for example Barnes et al. (2011) for more realistic models. The sensitivity of the emission profiles to the physical and dynamical state of hydrogen in and around galaxies makes isolating the intrinsic galaxy evolution from the IGM evolution very difficult.

Recent surveys probing beyond $z = 7$ have found a further complication: some observers have measured a luminosity dependence for the attenuation of quantities such as the luminosity function and the LAE fraction (Curtis-Lake et al., 2012; Konno et al., 2014; Zheng et al., 2017; Matthee et al., 2015; Santos et al., 2016). Faint ($M_{\rm UV} > -20.25$) LAEs are observed to decline in number in a similar manner beyond $z = 7$ as was seen for $z = 5 - 6$, and this has been used to extrapolate reionization histories. For bright ($M_{\rm UV} < -20.25$) LAEs however, a much slower evolution has been observed. This

can be most clearly seen in the luminosity functions of Zheng et al. (2017) and Ota et al. (2017), as well as the estimated LAE fraction in Stark et al. (2017). One suggested explanation (Zitrin et al., 2015) for this much weaker decline in the number of bright LAEs is that such galaxies sit in (and contribute ionising photons to) larger ionised bubbles, and hence are preferentially more visible than fainter galaxies.

There has been some recent theoretical work using simulations to explore the causes of these observations. Mason et al. (2018b) explored the effect of a mass-dependent intrinsic velocity offset in the emission profile of LAEs, finding that larger velocity offsets can increase the visibility of bright LAEs. Inoue et al. (2018) explored the effect of a mass-dependent optical depth in the host halo, and found such a dependence was required to explain observations. In this work we will further explore such effects, as well as the different roles the larger IGM environment can play around bright and faint LAEs.

There has also been some discussion in the literature of the effects of different selection techniques used for characterising LAEs (Stark et al., 2010), which can be divided into two categories: (i) (broadband) UV-selection with spectroscopic follow up (as in Stark et al., 2011, for example), and (ii) direct (narrowband) Lyα selection (as in Konno et al., 2014, for example). We note that observed LAE fractions are found via the former method, whilst most LAE luminosity functions are presented for populations found using the latter technique. In both cases the selection effects (such as AGN contamination) may play an important role in the inferred properties of high redshift LAEs. Importantly for our modelling, the selection technique will affect the mapping between galaxy mass and Lyα (or UV) luminosity. We discuss this further in Section 3.6.3.

In this Chapter we use the semi-analytic treatment of reionization combined with the high-dynamic-range Sherwood simulations (Bolton et al., 2017) as described in Section 2.2, to explore the effect of the IGM environment on the luminosity-dependent LAE evolution. In Section 3.2 we outline our simulation setup and calibration, which is based on Choudhury et al. (2015). In Section 3.3 we describe the framework we employ for calculating the transmission of Lyα radiation through the IGM. We establish models for reionization and for the LAEs in Section 3.4. We then present our results for these different models in Section 3.5. In Section 3.6 we discuss these results in comparison to other work, and finally draw conclusions in Section 3.7.

3.2 Simulation Method

In order to investigate the role of the IGM on LAE observations, we use cosmological hydrodynamical simulations with a semi-analytic treatment of reionization. There are two components to our numerical modelling: (i) a simulation of the (partially reionized) IGM, which includes the spatial distribution of neutral hydrogen, the peculiar velocities of the IGM gas and its temperature; (ii) a source model that produces galactic Lyα emission, which accounts for the spatial distribution of LAEs and their emission profiles.

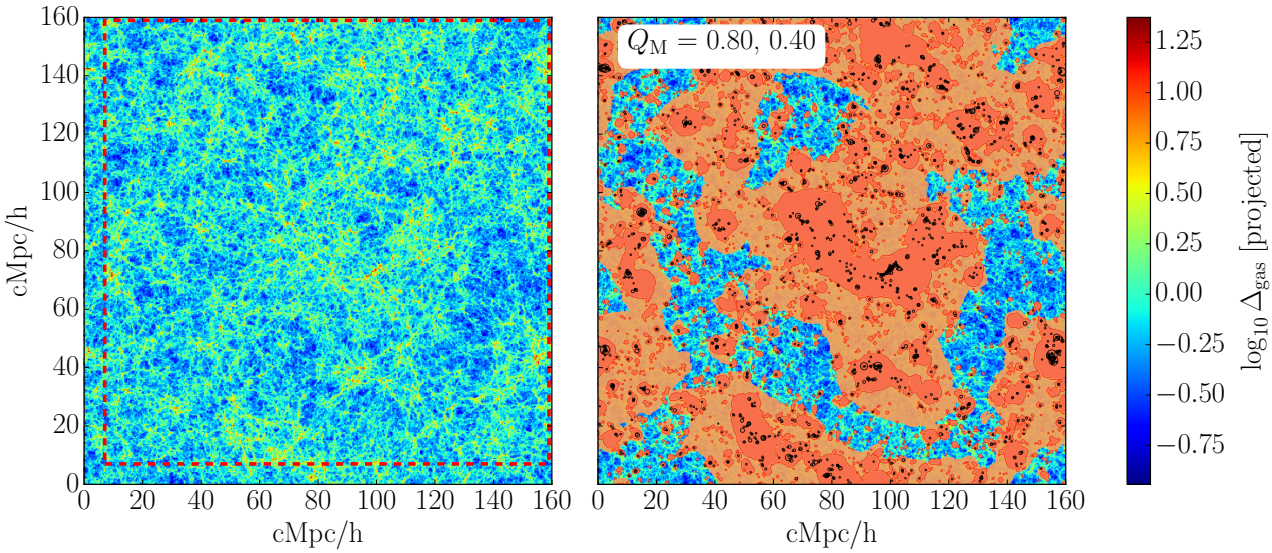


Fig. 3.1 Gas overdensity (left) and ionisation field (right, shaded) in a projected 0.5 cMpc/h slice of the $L = 160$, $N = 2048^3$, $z = 7$ snapshot from the Sherwood suite. The density has been interpolated onto a uniform 3D grid using the SPH kernel, and then a thin slice has been projected to create the 2D map. The dashed red square is a representative area for the comoving volume surveyed by Konno et al. (2018). For the ionisation structure on the right panel, the mass-averaged ionised fraction has been set to values of $\langle x_{\rm HII}\rangle_{\rm M} = 0.8$ and 0.4. The shaded areas mark ionised regions (found using the excursion set method described in Section 2.2.2) with the darker (brighter) shade showing the $\langle x_{\rm HII}\rangle_{\rm M} = 0.4$ (0.8) case. The positions of haloes from this slice are shown in black, with marker size proportional to halo mass.

For step (i), the simulation of the IGM, we follow the procedure in Choudhury et al. (2015), hereafter referred to as CPBH15, as described in Section 2.2. This approach starts from a cosmological hydrodynamic simulation, and then applies the excursion set formalism (Furlanetto et al., 2004c; Mesinger et al., 2011; Zahn et al., 2007) to determine the large-scale ionisation structure. We apply a self-shielding prescription that models the occurrence of neutral hydrogen embedded in ionised regions. Our reionization simulations are then calibrated to three different reionization histories, spanning the range consistent with CMB and Lyα forest data.

For step (ii), the source model, we also start with the same basic model used in CPBH15, aiming to reproduce their results. We then extend this basic model to try to account for differences between bright and faint LAEs. The details of these source models are outlined in Section 3.4.2.

3.2.1 Large-scale ionisation structure

The underlying cosmological hydrodynamical simulation used in this work is taken from the Sherwood simulation suite (Bolton et al., 2017), initially run as part of a PRACE simulation program. The simulation used for this work was performed in a box of length $L = 160$ cMpc/h[1] (where prefix c

[1]For test cases to compare with CPBH15 we also used a box of length $L = 80$ cMpc/h.

indicates *comoving* units, whilst prefix p indicates *proper* units). Snapshots of the initial PRACE run were saved for redshifts in the epoch of interest at $z = 6.0$, 7.0, 8.0 and 10.0. We have also re-run the simulation in order to better sample the EoR, saving snapshots every 40 Myrs; in particular in this work we make use of snapshots at $5.756 \leq z \leq 9.546$. For use in calculations the densities, velocities and temperatures of the particles were projected onto a grid using the SPH kernel. A projected slice of the density field from the simulation at redshift $z = 7$ can be seen in the left panel of Figure 3.1.

In CPBH15 a hybrid simulation was employed, comprised of an $L = 10$ cMpc/h, $N = 2 \times 512^3$ P-Gadget-3 hydrodynamical simulation to model the hydrogen distribution, and a larger low resolution dark matter-only simulation with $L = 100$ cMpc/h, $N = 1200^3$. These simulations were combined by tiling the small simulation box across the larger volume, making use of the ionisation structure and large-scale velocity modes of the large simulation box. We take advantage here of the much higher dynamic range of the Sherwood simulation suite and employ instead a single hydrodynamical simulation with almost twice the volume of their hybrid simulation, but at a factor two reduced spatial resolution compared to their 10 cMpc/h sized hydrodynamical simulation. Although lower in resolution this has the advantage of retaining the correlation between the gas density fields and the halo structure of the simulation, which was not present in the hybrid simulation of CPBH15. The larger volume also allows us to probe to higher halo masses, which is key to our modelling of bright and faint LAEs.

In recent observations, Konno et al. (2018) surveyed comoving volumes of $\sim 1.2 \times 10^7$ Mpc^3; our simulation volume ($\sim 1.3 \times 10^7$ Mpc^3) is therefore a better representation than the smaller volume of CPBH15 ($\sim 0.3 \times 10^7$ Mpc^3). In Figure 3.1 we show a representative survey area with a red dashed square for comparison with our box size.

We apply the reionization history modelling as described in Section 2.2, which includes the excursion set framework for determining ionised regions (discussed in Section 2.2.2) and the calibration of the background photoionisation rate (discussed in Section 2.2.4). We will return to our chosen reionization histories in Section 3.4. In the right hand panel of Figure 3.1 we show the ionisation field produced by this methodology for two different ionised fractions: in dark orange we show the ionised regions for $\langle x_{\rm HII} \rangle_{\rm M} = 0.4$, whilst the lighter orange region is at a higher fraction of $\langle x_{\rm HII} \rangle_{\rm M} = 0.8$. The positions of the haloes are overplotted as empty black circles, with the size of the marker proportional to the halo mass. As expected from the excursion set construction, the largest haloes sit in and dominate the largest ionised regions.

We note that the mean free path used in the calibration is fixed to the predicted values of a given reionization history model (see Section 3.4). We found this to be more stable than trying to calculate the mean free path iteratively from the simulation (using Eq. 2.28 and Eq. 2.22).

3.3 Lyman-α Transmission

Having performed the calibration as detailed in Section 2.2.4, we have simulation snapshots with realistic neutral hydrogen distributions that can be used to test the effect of the CGM and IGM on the transmission of the Lyα radiation from LAEs.

Early galaxies with high star-formation rates (SFRs) produce ionising radiation in their stellar component (Partridge & Peebles, 1967). This ionising radiation is then converted into Lyα line emission through recombination and collisional excitation of the gas in the ISM (Charlot & Fall, 1993; Dijkstra, 2014). The radiative transfer of Lyα photons through the ISM and CGM causes a diffusion in both physical and frequency space, resulting in a significant change to the emission profile. The photons that escape the galaxy must then traverse the IGM, which at $z > 6$ contains a significant non-zero neutral hydrogen fraction. Due to the resonant nature of Lyα absorption in neutral hydrogen, the presence of even small neutral fractions can alter the visibility of LAEs (see Meiksin, 2009; Dijkstra, 2014, for reviews of IGM and Lyα physics).

As discussed in Section 3.1, observations of LAEs at high redshifts have found a decline in number densities. Explaining these observations is made difficult by the degeneracy between internal galaxy evolution (parametrised by the fraction of Lyα photons that escape galaxies, $f_{\mathrm{esc,Ly}\alpha}$, which may be a function of z) and IGM absorption (parametrised by the neutral fraction x_{HI}) (Dayal et al., 2009). In this work we consider the effect of the CGM/IGM only, and do not model galaxy evolution.

3.3.1 Lyα transmission fraction

In order to quantify the effect of the IGM and CGM on the transmission of Lyα photons, we extract sightlines from our simulation snapshots that pass through LAE host haloes, and calculate the radiative transfer along them (see Section 3.4 for details on how LAE host haloes are selected). The sightlines are chosen to be 160 cMpc/h in length, parallel to the simulation box axes. We take advantage of the periodic boundary conditions of the simulation to translate the start of the sightline such that the halo is positioned at the centre[2]. The gas properties are initially gridded into 2048 bins (78.13 ckpc/h cell size), with a further 2048 bins in a high resolution region of length 20 cMpc/h (giving a 9.77 ckpc/h cell size) containing the host halo[3]. This ensures we resolve the gas around the host halo, including small-scale high density regions likely to self-shield.

Neglecting scattering by dust, the equation of radiative transfer can be written (Draine, 2011),

$$J_\nu(\tau_\nu) = J_\nu(0)e^{-\tau_\nu} + \int_0^{\tau_\nu} d\tau'_\nu \, e^{-(\tau_\nu - \tau'_\nu)} \, S_\nu(\tau'_\nu), \tag{3.1}$$

$$\approx J_\nu(0)e^{-\tau_\nu}, \tag{3.2}$$

[2]The optical depth calculation was found to converge on considerably smaller spatial scales than 80 cMpc/h.
[3]We note that the softening length used in these simulations is 3.13 ckpc/h.

where $J(\nu)$ is the galaxy emission profile (the specific intensity of radiation at frequency ν), $\tau(\nu)$ is the Lyα optical depth (see Section 3.3.2 below), and the source function, S_ν, is approximately zero because the Lyα emissivity of the IGM gas is negligible (Silva et al., 2016). This expression allows us to calculate the emission profile of a galaxy after re-processing by the surrounding IGM gas, $J'(\nu) = J(\nu)e^{-\tau(\nu)}$. With this we can calculate the transmission fraction of photons (or transmissivity) given by (Mesinger et al., 2015),

$$T^{\rm IGM}_{{\rm Ly}\alpha} = \frac{\int_{\nu_{\rm min}}^{\nu_{\rm max}} d\nu\, J(\nu)\, e^{-\tau(\nu)}}{\int_{\nu_{\rm min}}^{\nu_{\rm max}} d\nu\, J(\nu)}, \tag{3.3}$$

where $J(\nu)$ is appropriately normalised. Since we place the Lyα emitter at the centre of the sightline, the frequency limits in Eq. (3.3) are the Lyα frequency blue/redshifted along half the sightline length, which extends considerably beyond the wings of the emission profile.

We note that this "$e^{-\tau}$ modelling" of the radiative transfer of Lyα photons has been compared to full radiative transfer by Zheng et al. (2010). They suggested that such models can over-attenuate the line profile compared to that of full calculations because some of the frequency diffusion is neglected. A balance has to be struck between the frequency diffusion in the inner parts of the galaxy and the attenuation by the neutral hydrogen surrounding the galaxy. We will account for the frequency diffusion in the inner part of the galaxies in our modelling of the spectral distribution (see Section 3.4).

3.3.2 Lyα attenuation due to the CGM and IGM

As suggested in Dijkstra (2014) we split the Lyα optical depth responsible for attenuating the Lyα emission from galaxies into two contributions: (i) $\tau_{\rm HI}(z, v)$, the opacity due to any recombined neutral hydrogen or self-shielded regions *within* ionised bubbles; (ii) $\tau_{\rm D}(z, v)$, the opacity due to damping-wing absorption in the residual neutral IGM. Note that these quantities depend on the velocity offset, v, which is determined by both the Hubble flow and the difference in peculiar velocity of emitter and absorber. So we can calculate,

$$\tau_{{\rm Ly}\alpha}(v) = \tau_{\rm HI}(v) + \tau_{\rm D}(v). \tag{3.4}$$

Physically, photons emitted close to line centre will redshift out of resonance as they traverse the IGM. It is important to consider that scattering/absorption occurs at velocity shifts close to zero in the absorber's rest frame. This means that redshifted photons in the frame of neutral gas infalling onto the host halo can be blueshifted back into resonance.

We note that both of the components in Eq. (3.4) are calculated in the same manner. In order to calculate the optical depth we assume a Voigt profile for the absorption cross section, in particular using the analytic approximation from Tepper-García (2006),

$$H(a, x) = e^{-x^2} - \frac{a}{\sqrt{\pi}x^2}[e^{-2x^2}(4x^4 + 7x^2 + 4 + 1.5x^{-2}) - 1.5x^{-2} - 1], \tag{3.5}$$

where $H(a, x)$ is the Hjerting function, related to the Voigt profile by (Rybicki & Lightman, 1985),

$$\phi(\nu) = (\Delta\nu_{\rm D})^{-1}\pi^{-1/2}H(a, x), \tag{3.6}$$

$$\text{where } \Delta\nu_{\rm D} \equiv \frac{\nu_\alpha}{c}\sqrt{\frac{2k_BT}{m_{\rm H}}}, \tag{3.7}$$

$$a \equiv \frac{\Lambda_\alpha}{4\pi\Delta\nu_{\rm D}}, \tag{3.8}$$

$$x \equiv \frac{\nu - \nu_\alpha}{\Delta\nu_{\rm D}}. \tag{3.9}$$

Note in the above formulae we have used: the Lyα frequency $\nu_\alpha = 2.46 \times 10^{15}$Hz, the hydrogen $2p \longrightarrow 1s$ decay rate $\Lambda_\alpha = 6.25 \times 10^8 \text{s}^{-1}$, the Boltzmann constant k_B, the hydrogen atomic mass $m_{\rm H}$ and the temperature of the gas, T, at the absorber. For a given sightline, we find the optical depth in a (redshift-space) cell $\mathbf{i}$ by summing up all the contributions from positions in front of the emitter (Bolton & Haehnelt, 2007), where we define $v = 0$ at the position of the emitter,

$$\tau(\mathbf{i}) = \frac{\nu_\alpha\sigma_\alpha\delta R}{\sqrt{\pi}}\sum_j^N \frac{n_{\rm HI}(j)}{\Delta\nu_{\rm D}(j)}H(a, x(i, j)), \tag{3.10}$$

where δR is the cell width, and the cell has Hubble velocity $v_{\rm H}(\mathbf{i})$ and peculiar velocity $v_{\rm pec}(\mathbf{i})$, such that

$$x(i, j) = \sqrt{\frac{m_{\rm H}}{2k_BT}}\,[v_{\rm H}(\mathbf{i}) - v_{\rm H}(\mathbf{j}) - v_{\rm pec}(\mathbf{j})]. \tag{3.11}$$

Eq. (3.10) is the optical depth to Lyα emission from the halo position, which is then redshifted along the sightline. In velocity space absorption can appear to occur "behind" the halo due to the non-negligible width of the absorption profile, and because of infalling matter around the halo.

3.3.3 Lyα scattering in the host halo

The importance of carefully modelling the neutral gas in and close to the host halo was emphasised and explored in Sadoun et al. (2017), hereafter referred to as SZM17. In Figure 3.2 we show spherically averaged density profiles for three mass bins spanning the masses of the halo population in our simulations. The neutral hydrogen densities were calculated assuming a fixed value of $\Gamma_{\rm HI}$ and solving for photoionisation equilibrium using Eq. (2.21), including the self-shielding prescription discussed in Section 2.2.5. Note however that spherically averaging will smooth out the overdensities surrounding the halo which are used to calculate the amount of self-shielding; this means that these radial profiles somewhat under-represent the neutral gas density compared to sightlines through our simulations which are not spherically averaged. We see more extended profiles in the more massive haloes,

whereas in the less massive halos the profiles are more peaked around the central halo position. In the central panel we also compare to profiles presented for haloes with mass $M_h = 10^{10.5}\ M_\odot$ by SZM17. We note that the total hydrogen density profiles are similar for $r > 20$ pkpc, however at smaller radii both our total and neutral hydrogen densities are lower than the model presented by SZM17. This is likely due to the QUICK_LYALPHA star formation prescription, which converts dense gas into star particles (as described in Section 2.1.5). This prescription will therefore remove some of the very dense gas in the centres of haloes, as we see in Figure 3.2.

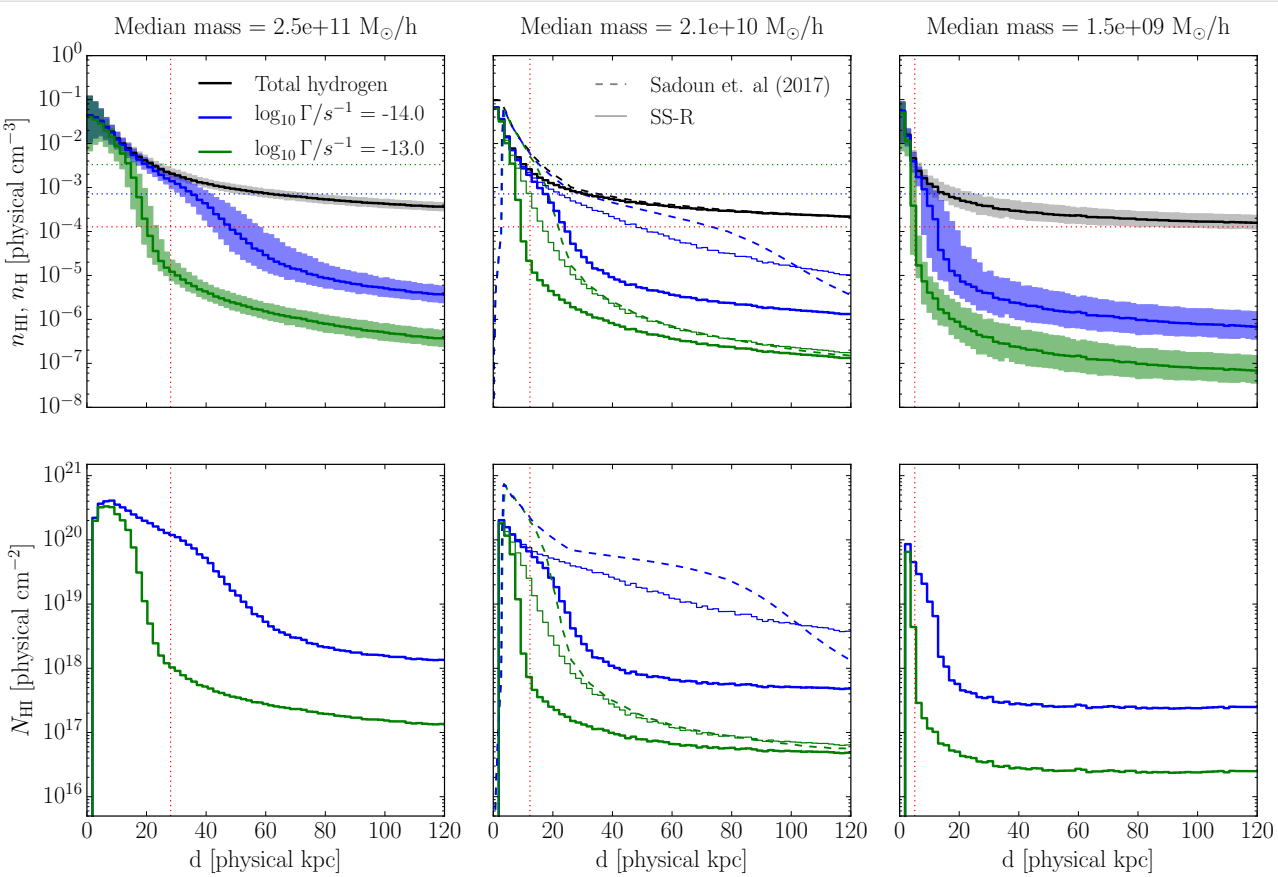


Fig. 3.2 *Top panels*: Median hydrogen number density profiles at $z = 7$, spherically averaged, for both total hydrogen in black, and neutral hydrogen in green ($\Gamma = 10^{-13}$ s^{-1}) and blue ($\Gamma = 10^{-14}$ s^{-1}). The three panels correspond to samples of 500 haloes with mass bins that bracket the halo population of our simulations; most massive on the left, least massive on the right, and a sample chosen for comparison with SZM17's $M_h = 10^{10.5}\ M_\odot$ model in the middle panel. The red vertical dotted line shows the median virial radii, whilst the horizontal red line shows the mean baryonic density. The horizontal green and blue dotted lines show the self-shielding density thresholds for the different photoionisation rates. In the middle panel we also show the profiles from SZM17 with dashed lines, and the profiles found using the Rahmati et al. (2013) self-shielding prescription (labelled SS-R) shown with thin stepped lines. The shaded regions indicate 68% scatter in the samples (not shown in the middle panel to aid visual comparison with SZM17). *Bottom panels*: Column densities at a given radius estimated by multiplying the number density by the radial distance.

We also show the neutral hydrogen profiles for the widely used self-shielding prescription proposed by Rahmati et al. (2013) with the thin (step) curves. These are closer to those presented by SZM17, especially for the lower photoionisation rate of $\Gamma_{\rm HI} = 10^{-14}{\rm s}^{-1}$. We suggest that most of the difference between the profiles in our simulations and the modelling of SZM17 is due to the presence of ionising sources. In the simulations on which the prescription of Chardin et al. (2018) is based, there are ionising sources within the self-shielded regions which affect the local photoionisation rate and therefore the self-shielding threshold density. These are not accounted for in the Rahmati et al. (2013) prescription. Note that while SZM17 do account for a central ionising source in their calculation, they assume this source to be rather weak. As already mentioned some of the difference will also be due to the spherical averaging which is not accounted for in our self-shielding prescription. Note further that in this work we also consider the role of the larger scale ionisation structure, and the presence of an IGM volume-filling neutral fraction, which SZM17 neglect. As discussed by SZM17, the attenuation near to the host halo is very sensitive to the distribution of neutral hydrogen close to the Lyα emitters. We discuss this in more detail in Section 3.6.2.

CPBH15 and Bolton & Haehnelt (2013) did not attempt to simulate the complex radiative transfer within the host halo, but instead assumed an intrinsic emission profile (for photons leaving the host system, but before attenuation by the IGM) and argued that this accounts for these effects. In those works the contributions of neutral gas within 20 pkpc were therefore neglected around the halo; for the narrower range of halo masses considered in those works this was a consistent and sufficient exclusion. Our modelling here includes a considerably larger range of halo masses, which therefore also have a considerable range of virial radii. Excluding gas within a fixed distance of 20 pkpc uniformly across our halo population would remove all the neutral gas within a few virial radii around the less massive haloes, whilst only remove the gas within a fraction of the virial radius in the most massive haloes. Here we therefore choose the exclusion region based on the mass (or virial radius) of the host halo and will use our simulation and the $e^{-\tau}$ approach to account for the attenuation due to the neutral hydrogen in the outer part of the host haloes of Lyα emitters.

We have tested the effect of varying the size of the exclusion region by excluding gas within 0.5, 1.0, 2.0, 5.0, 10.0 $R_{\rm vir}$, where we use $R_{\rm vir} = R_{\rm 200,crit}$. The resulting transmission curves for these exclusions, calculated as described in Section 3.3.2, are shown in Figure 3.3. In the left-hand panels in shades of blue we present the results for a sample of less massive haloes, whilst in the right-hand panels in shades of red we show the results for more massive haloes. The important role of the gas peculiar velocities can be seen in the top panels by comparing the solid lines (full calculations) to the dashed lines (calculated neglecting peculiar velocities). In particular in the more massive haloes, the peculiar velocities are sufficient to dramatically move the position of the damping wing. We also note, considering the solid lines, that the more massive haloes are more sensitive to the choice of exclusion: in the less massive haloes (blue lines) the damping wing of the profile is moved by ~ 150 km/s between the two exclusion extremes shown, whilst in the more massive haloes it is moved by ~ 350 km/s.

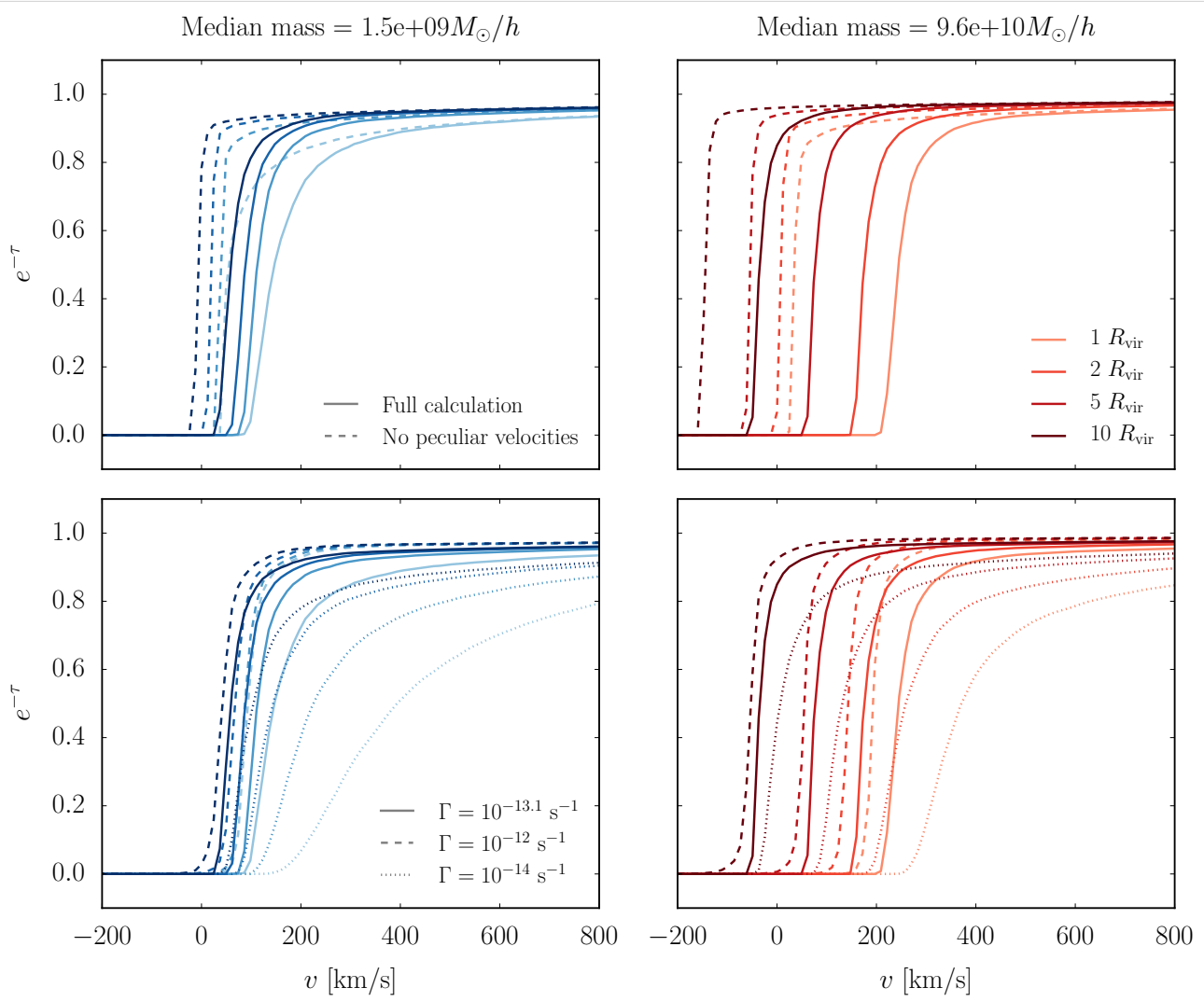


Fig. 3.3 Median transmission curves, testing different exclusion radii (indicated by darkness of line), for a less massive sample of haloes (left panels) and a more massive sample (right panels). In the top panels the solid lines show the results from the full calculation, whilst the dashed lines show what happens when peculiar velocities are neglected. Both of these panels assume $\log_{10}\Gamma/\mathrm{s}^{-1}$= -13.1. The bottom panels compare three different photoionisation rates: $\log_{10}\Gamma/\mathrm{s}^{-1}$= -12 (dashed), -13.1 (solid), -14 (dotted).

In the bottom panels of Figure 3.3 we show the effect of varying the chosen background photoionisation rate Γ. This leads to a change in the amount of equilibrium neutral hydrogen (self-shielded or recombined) within ionised regions close to the halo. We see that for the higher photoionisation rate the effect of changing the exclusion region is reduced, and vice versa for the lower photoionisation rate.

Our fiducial choice is to exclude gas within 1.0 $R_{\rm vir}$; unless otherwise specified, all results presented hereafter were calculated with this choice. As can be seen in Figure 3.3, there will be some dependence of the Lyα transmission on the chosen exclusion region. We mitigate this dependence with our

choice of source models, as detailed in Section 3.4. Further consequences of our choice of the size of the exclusion region are discussed in Section 3.6.2.

3.3.4 Transmission fraction ratios (TFRs)

As we are primarily interested in the evolution of the Lyα attenuation during the epoch of reionization we consider the ratio of transmission fractions $T(z)/T(z_{\rm ref})$ (hereafter referred to as TFRs), where $z_{\rm ref}$ is a reference redshift. In particular we choose to construct the ratio of higher redshifts with respect to $z_{\rm ref} = 5.756$, matching the choice of $z = 5.7$ common in the literature.

Narrowband (Lyα-selected) observations of LAEs at different redshifts can be used to calculate the TFR evolution as (Konno et al., 2018),

$$\frac{\overline{T}(z)}{\overline{T}(z_{\rm ref})} = \frac{\kappa(z_{\rm ref})}{\kappa(z)} \frac{f_{{\rm esc,Ly}\alpha}(z_{\rm ref})}{f_{{\rm esc,Ly}\alpha}(z)} \frac{\rho_{{\rm Ly}\alpha}(z)/\rho_{{\rm Ly}\alpha}(z_{\rm ref})}{\rho_{\rm UV}(z)/\rho_{\rm UV}(z_{\rm ref})}, \tag{3.12}$$

where $L_{{\rm Ly}\alpha} = \kappa L_{\rm UV}$, $f_{{\rm esc,Ly}\alpha}$ is the escape fraction of Lyα photons, and $\rho_{\rm UV}$ is the intrinsic UV luminosity density whilst $\rho_{{\rm Ly}\alpha}$ is the observed (attenuated) Lyα luminosity density. This relative transmission fraction is an effective way of quantifying the evolution observed in the LAE luminosity function. In particular it is a convenient quantity that allows one to estimate the neutral fraction $x_{\rm HI}$ from an observational sample. In this work we also choose to calculate the TFR evolution rather than the luminosity function evolution because it can be calculated via Eq. (3.3) independently of the uncertain relationship between the LAE host halo's mass and its Lyα luminosity. We leave the explicit modelling of the M_h-$L_{{\rm Ly}\alpha}$ relation, and hence the luminosity function evolution, to Chapter 4.

3.3.5 Lyα Fractions

Alongside the evolution of the Lyα luminosity function, observers have also measured the evolution of the fraction of continuum-selected galaxies which emit strongly in Lyα. This is determined using samples of UV-selected galaxies (via the Lyman break technique), with follow-up spectroscopy to measure Lyα. The fraction, $X_{{\rm Ly}\alpha}$, is the proportion of such an LBG sample that are measured to have a Lyα equivalent width above a given threshold (Stark et al., 2011; Ono et al., 2012; Treu et al., 2012)

In this work we also calculate the predicted evolution of $X_{{\rm Ly}\alpha}$ following a similar strategy to Sadoun et al. (2017) and CPBH15. We start with the presription of Dijkstra et al. (2011) in which we derive the *rest-frame equivalent width* (REW) distribution. This is done by assuming that there is a probability distribution $P_{\rm int}(>{\rm REW})$ for an intrinsic unabsorbed REW distribution which does not evolve with redshift; the observed redshift evolution is then entirely due to the attenuation by the IGM. Given the probability distribution for the transmitted fractions at a given redshift $P_T(T, z)$ and the intrinsic distribution, we can find the REW distribution at that redshift as,

$$P(>{\rm REW}, z) = \int_0^1 {\rm d}T\ P_T(T, z)\ P_{\rm int}(>{\rm REW}/T). \tag{3.13}$$

As in CPBH15 we choose to determine $P_{\rm int}(>\mathrm{REW})$ as the function which gives $P(>\mathrm{REW}, z=6)$ that matches the observational data of Stark et al. (2011). We fit the following functional form for the intrinsic distribution (Shapley et al., 2003),

$$P_{\rm int}(>\mathrm{REW}) = \exp(-\mathrm{REW}/\mathrm{REW}_0)/(\mathrm{REW}_0 + \mathrm{REW}_1), \tag{3.14}$$

where REW_0 and REW_1 are free parameters which vary depending on the simulated transmission fraction distribution. Given this intrinsic distribution, and using Eq. (3.13), we can find the fraction of Lyα emitting galaxies over a given threshold equivalent width as,

$$X_{\mathrm{Ly}\alpha}(\mathrm{REW}, z) = P(>\mathrm{REW}, z). \tag{3.15}$$

The values predicted by the simulations can then be compared to observed fractions.

3.4 Models

Using the above simulation setup and Lyα transmission framework, we can explore different models of reionization and LAEs to compare with current observations. In particular we test three reionization histories which bracket the possible progress of reionization at a given redshift. We also employ three different models for the masses of the host haloes of LAEs to explore the effect of host halo mass on Lyα transmission. We therefore test a total of nine possible model combinations.

3.4.1 Reionization Histories

We consider here three different reionization histories first discussed in CPBH15; we follow the naming convention established in Kulkarni et al. (2016). As outlined in section 3.2.1, each model provides $\langle x_{\rm HII}(z)\rangle_M$ and $\lambda_{\rm mfp}(z)$ which we input into the calibration calculation.

- **HM12**: this ionisation history corresponds to the commonly used model of Haardt & Madau (2012), based on the meta-galactic UV background. We use $\langle x_{\rm HII}(z)\rangle_M$ and $\lambda_{\rm mfp}(z)$ as predicted in Haardt & Madau (2012). In this model the galactic UV emission is used as a tracer of the cosmic star formation history; this can be derived from the galaxy UV luminosity function (Robertson et al., 2013). Importantly the main contribution to the ionising photon budget comes from galaxies, with quasars and early Population III stars playing a negligible role. The universe is completely ionised in this model by $z = 6.7$. Comparing the model predictions to observed data, it agrees reasonably well with observed background photoionisation rates (Faucher-Giguère et al., 2009; Calverley et al., 2011; Wyithe & Bolton, 2011). However its prediction for the Thomson optical depth of the CMB, $\tau_{\rm el} = 0.084$, is higher than the measurement of Planck Collaboration (2016) by more than 1 σ.

- **Late**: this model uses the same evolution as the HM12 model with $\langle x_{\rm HII}(z)\rangle$ shifted in z such that reionization completes at $z = 6$ instead of $z = 6.7$, but with the same $\mathrm{d}\langle x_{\rm HII}\rangle_{\rm M}/\mathrm{d}z$. A similar reionization history was found in the full radiative transfer simulations of Chardin et al. (2015), hereafter referred to as Ch15. In Ch15 the radiative transfer code Aton (Aubert & Teyssier, 2008) was used to post-process high resolution cosmological hydrodynamical simulations calibrated to Lyα forest data in order to calculate the evolution of the ionising photon mean free path. We use the mean free path predicted in that work for our calibration. The CMB Thomson optical depth is $\tau_{\rm el} = 0.068$ in this model.

- **Very Late**: Reionization completes at $z = 6$ as in the Late model, but the evolution of $\langle x_{\rm HII}\rangle_{\rm M}$ is much more rapid for $z > 6$. AGN dominated reionization could lead to the history that this model predicts, see Kulkarni et al. (2017) for further details. We predict the mean free path for this model using the relationship between $\langle x_{\rm HII}\rangle_{\rm M}$ and $\lambda_{\rm mfp}$ from the Late model[4]. The Thomson optical depth in this case is $\tau_{\rm el} = 0.055$.

We follow Kulkarni et al. (2016) in choosing the Late reionization history as our fiducial model.

In Figure 3.4 we show the final calibrated parameters of the simulation, including the reionization histories for $\langle x_{\rm HII}(z)\rangle_{\rm M}$. The HM12, Late and Very Late calibrated parameters are shown as blue circles, red triangles and grey inverted triangles, respectively, in all panels. The solid black lines in all panels show the predictions of the underlying model from HM12 (Haardt & Madau, 2012), whilst the red dashed lines show the predictions from Ch15 (Chardin et al., 2015). The fixed quantities are $\langle x_{\rm HII}(z)\rangle_{\rm M}$ and $\lambda_{\rm mfp}(z)$, shown on the left-most panels. We see reionization progresses from high redshift (where $\langle x_{\rm HII}\rangle_{\rm M} \longrightarrow 0$) until around $z \sim 6$; specifically in the HM12 model we see $\langle x_{\rm HII}\rangle_{\rm M} = 1$ at $z = 6.7$, whilst in the other models it reaches 1 at $z = 6$. The optical depth of the CMB to electron scattering is predicted by the reionization history models, shown in the top middle panel. Here the three lines for each model can be compared to the Planck Collaboration (2016) value shown as a horizontal green line, with green shading indicating the 1 σ bounds. The quantities derived during our self-consistent calibration are the clumping factor (bottom middle panel), the ionising emissivity (bottom right panel) and the background photoionisation rate (top right panel). We see in the HM12 model that the mean free path and the photoionisation rate increase at a largely constant exponential rate as reionization progresses, with a roughly constant ionising emissivity. This smooth evolution of the mean free path may however be unrealistic (Puchwein et al., 2019). In comparison, the Late and Very Late models predict a more steady photoionisation rate at high redshifts, which suddenly increases close to percolation at $z \sim 7$ when the H II regions overlap to an extent that the mean free path of ionising photons rises rapidly. For this more abrupt end to reionization to occur there needs to be a sharper increase in the mean free path, which can be seen in the bottom left panel. We note that the recent physically-motivated model of Puchwein et al. (2019) has been able to reproduce this required rather sharp increase.

[4]Beyond $z = 6$, both of these quantities are monotonically increasing with redshift, and hence can be mapped together. This allows us to find the mean free path for a given $\langle x_{\rm HII}\rangle_{\rm M}$ of the Very Late model.

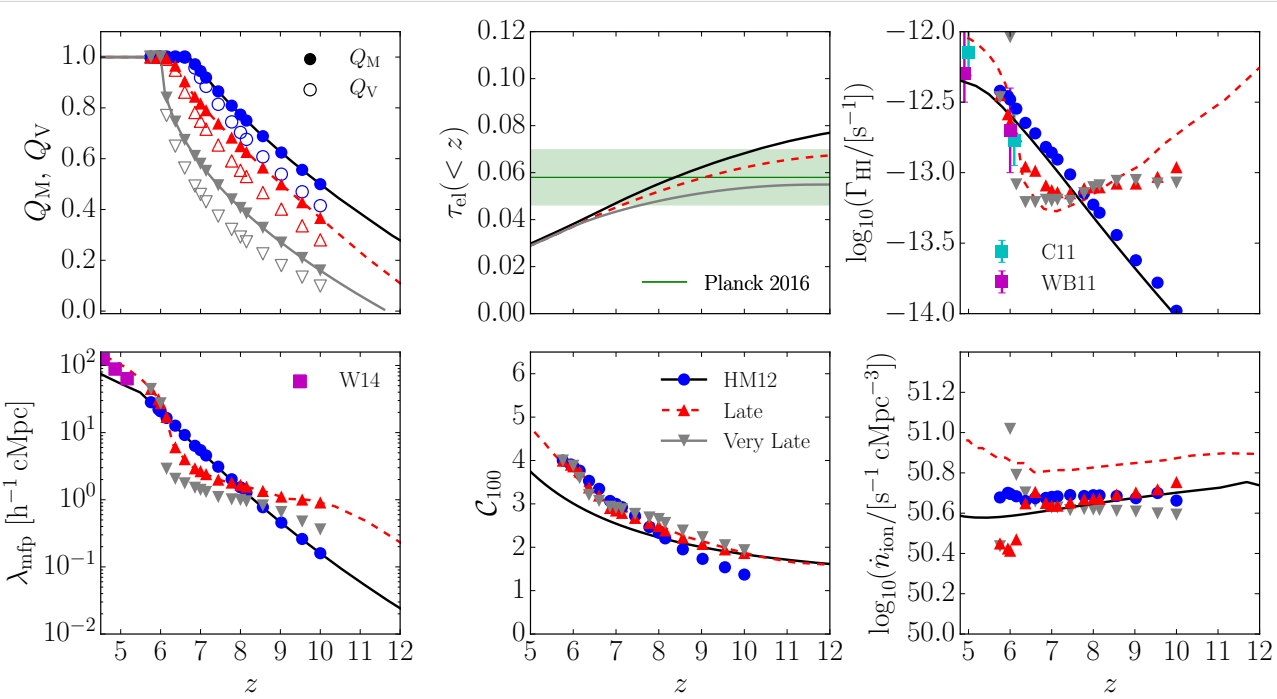


Fig. 3.4 Calibrated parameters of the simulation. Clockwise from the top left: mass- and volume-averaged ionised fractions $Q = \langle x_{HII} \rangle$, electron-scattering optical depth, UV background photoionisation rate, ionising emissivity, clumping factor for overdensities less than 100, ionising photon mean free path within ionised regions. Our chosen models are shown as blue dots (HM12), red triangles (Late) and grey inverted triangles (Very Late). The reionization histories are shown as solid black (Haardt & Madau, 2012, HM12) and red dashed (Chardin et al., 2015, Ch15) lines. Observed data from Calverley et al. (2011, C11), Wyithe & Bolton (2011, WB11), Worseck et al. (2014, W14) are overplotted for comparison. Note that in the top middle panel (electron-scattering optical depth) we show the 1 σ bounds shaded in green from Planck Collaboration (2016)

3.4.2 Host halo masses

To model the effect of the IGM and CGM on the Lyα emission, we have to simulate the underlying signal from the galaxies. This step of our simulation has two components: (i) the spatial distribution of galaxies in our simulation volume; (ii) the emission profile, $J(v)$, of the galaxies. We expect the galaxy spatial distribution to follow the halo distribution (Kaiser, 1984; Verde et al., 2002). Unfortunately the emission profile for high redshift galaxies is poorly constrained. Our modelling choices are motivated by the tests discussed in Section 3.3.3.

We consider three models for the spatial distribution of LAEs, based on different halo mass bins, choosing a sample of 4992 haloes per model. These models therefore have varying levels of correlation between the LAE positions and the positions/size of ionised regions.

- **Small mass**: firstly we place the LAEs in haloes smaller than the mean mass, which on average have a mass $M_h \sim 10^9$ $M_\odot$/h. This simple model is useful for understanding the evolution of faint LAEs.

- **Large mass**: secondly we consider the case where LAE positions have maximal correlation with the ionised regions, by placing them in the most massive haloes of the simulation volume. These haloes have masses in the range $10^{11} \lesssim M_h \lesssim 10^{12}$ $M_\odot$/h. This model is used to represent the bright end of the LAE distribution.

- **Continuous**: finally we place LAEs in a random sample taken from the full halo population of the simulation, noting that the mass resolution of the simulation naturally enforces a physically realistic cutoff mass $M_h > 10^7 M_\odot$ (Finlator et al., 2017). Due to the steep slope of the halo mass function, this model will be dominated by smaller more common haloes, and hence will be similar in many respects to the small mass model. It is intended as middle ground between the first two models, and we consider it the most realistic model for comparing with an observational survey of average LAEs.[5]

The first two models are used as approximate representations of the different populations of faint (lower mass host haloes) and bright (higher mass host haloes) LAEs.

Table 3.1 Averages masses of the different halo mass models used in this chapter.

Name	$\log_{10}(\overline{M}_{\rm h}$ [$M_\odot$])			
	$z = 6$	$z = 7$	$z = 8$	$z = 10$
`Small mass`	9.393	9.358	9.328	9.283
`Large mass`	11.531	11.259	11.002	10.518
`Continuous`	9.594	9.512	9.477	9.370

In Figure 3.5 we show the median velocity for the gas distribution along sightlines through the small mass (cyan lines) and large mass (magenta lines) haloes. The figure shows much larger infalling velocities around the large mass haloes. Comparing across the different redshifts (with $z = 10$ represented by the dash dotted lines, up to $z = 6$ represented by the solid lines) we also see more significant evolution in the larger mass haloes than for the smaller mass haloes. This evolution is largely driven by the evolution in the halo masses of our large mass model, which can be seen in Table 3.1. Therefore our large mass model represents an upper limit on the possible contribution the local gas environment evolution can provide towards Lyα attenuation.

We note that the peculiar velocities tend to zero with increasing radius, but only on large scales of order 80 cMpc/h. As a result of the long-range correlations of peculiar velocities, out to large radii from the host halo the gas is infalling with respect to the halo. Comparing to the neutral gas density profiles in Figure 3.2 we see that the high column density gas around the more massive haloes will be moved towards line centre (in the gas rest frame) by the large infalling velocities. Comparing

[5] Although we do not need to explicitly specify a mass-luminosity mapping for the results in this chapter, we note that for the commonly assumed linear relation of Lyα luminosity and host halo mass, $L_{\rm Ly\alpha} \propto M_h$, the continuous model would correspond to a random sampling of the faint end of the luminosity function.

these velocity profiles to the transmission curves in Figure 3.3 suggests that there can be increased attenuation due to damping wing absorption by the neutral (self-shielded) gas around massive haloes compared to the less massive haloes of the small mass model.

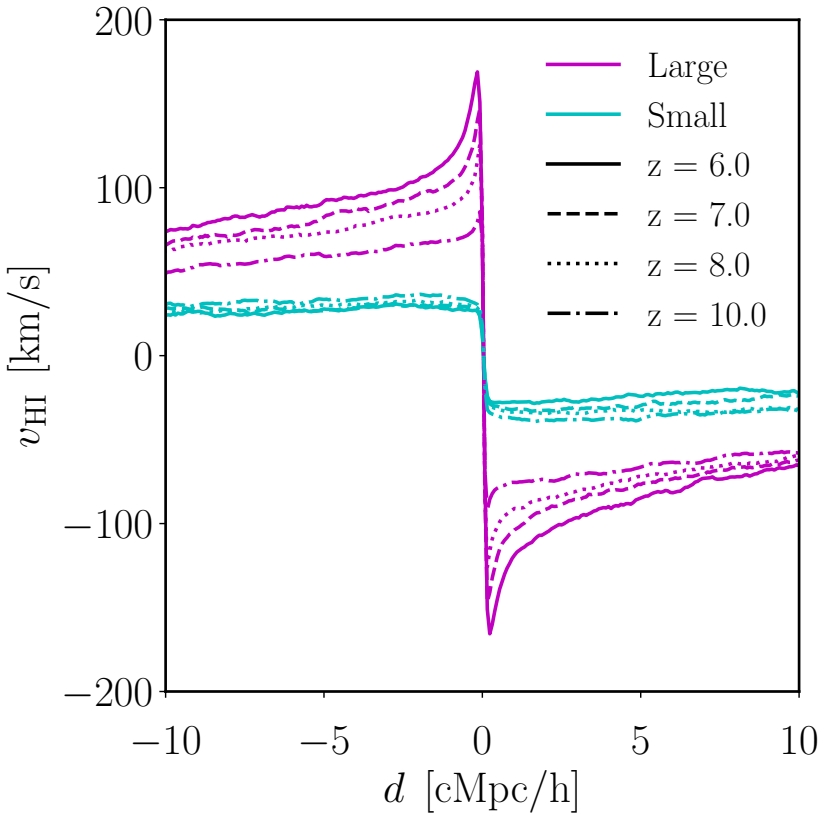


Fig. 3.5 Median values of the hydrogen peculiar velocity around haloes (at $d = 0$ [cMpc/h]) are shown for the small mass range (cyan) and large mass range (magenta) for 5000 sightlines at $z = 6$ (solid), $z = 7$ (dashed), $z = 8$ (dotted) and $z = 10$ (dash-dotted).

For the second component of our source model we assume a Gaussian emission profile, with centre offset (in the galaxy rest-frame) from Lyα by a shift $\Delta v = v_\alpha \Delta v_{\rm int}/c$, and width given by $\sigma_v = v_\alpha \sigma_v/c$. Importantly we account for the peculiar velocity of the emitter when using the emission profile for calculations in the frame of the sightline. The radiative transfer through the ISM produces a characteristic double-peaked emission profile (Dijkstra, 2014), however the blue peak will redshift into resonance while the photons traverse the IGM. At the considered redshifts even residual neutral gas in ionised regions is sufficient to render this blue peak unobservable, hence our use of a singly-peaked Gaussian emission profile. It has been empirically established that the Lyα emission line-centre is offset in both high redshift LAEs and lower-redshift analogs (Stark et al., 2015; Erb et al., 2014). A suggested explanation for the cause of this offset is galactic outflows (Steidel et al., 2010; Shibuya et al., 2014), but almost certainly in combination with resonant scattering effects (e.g. Barnes et al.,

2011). We use the same values of $\Delta v_{\rm int}$ and σ_v that were employed as the default model of CPBH15. The emission profile is the same for all the haloes, with $(\Delta v_{\rm int}, \sigma_v) = (100, 88)$ km s^{-1}. These values are similar to those inferred in Stark et al. (2015) using the C [III]λ1909 line.

In summary we have nine model permutations, which include the three reionization histories and the three halo mass models. Our fiducial model for comparison with observational data is the 'Late' reionization history combined with the continuous mass model.

3.4.3 Observational constraints on host halo masses from LAE clustering

The best constraint on host halo masses of LAEs can be obtained using clustering statistics. The estimates for $z = 6.6$ LAEs from Ouchi et al. (2010); Sobacchi & Mesinger (2015); Ouchi et al. (2018) are shown in Table 3.2. The average masses of host haloes have been calculated in the above works using samples that span the luminosity range from faint ($10^{42} \lesssim L_{{\rm Ly}\alpha} < 10^{43}$ erg/s) to bright ($L_{{\rm Ly}\alpha} \gtrsim 10^{43}$ erg/s), and so do not necessarily reflect the expected masses for this distinction, but rather an average of the two ranges. We leave it to future work to perform a detailed clustering analysis on the observed samples of LAEs split into these luminosity brackets. For comparison with this work, the average host halo masses at representative redshifts for our small and large mass models are shown in Table 3.1.

Table 3.2 Estimated (observed) average host halo masses at $z = 6.6$, using clustering statistics such as the angular correlation function (ACF).

Work	$\log_{10}(\overline{M}_{\rm h}\,/[{\rm M}_\odot])$
Ouchi et al. (2010)	10–11
Sobacchi & Mesinger (2015)	$\lesssim 10$
Ouchi et al. (2018)	$10.8^{+0.3}_{-0.5}$

Note again the definition of our mass models: *large* corresponds to the most massive haloes in the simulation, which evolves with redshift; *small* corresponds to the most common haloes with mass $\sim 10^9$ M$_\odot$. A comparison of Tables 3.2 and 3.1 shows that the observed masses lie somewhere in between our small and large mass models. As mentioned in Section 3.4.2 our *continuous* model should thus be the most representative of a real LAE sample. Although the steepness of the halo mass function biases the average mass towards the smaller mass end of the spectrum, we still expect there to be LAEs hosted by the more massive haloes considered here.

3.5 Results

Having applied our calibration scheme for the different reionization history models, and then calculated the Lyα transmission for the different LAE models, we can now explore the effect of these different model parameters on the distribution of transmission fractions. We can also explore the

effect on the TFR (transmission fraction ratio, as defined in Section 3.3.4) evolution, and compare this to the observed difference between bright and faint LAEs. Finally we can also derive the evolution of the Lyα fraction, $X_{\mathrm{Ly}\alpha}$, and compare our predictions with observations.

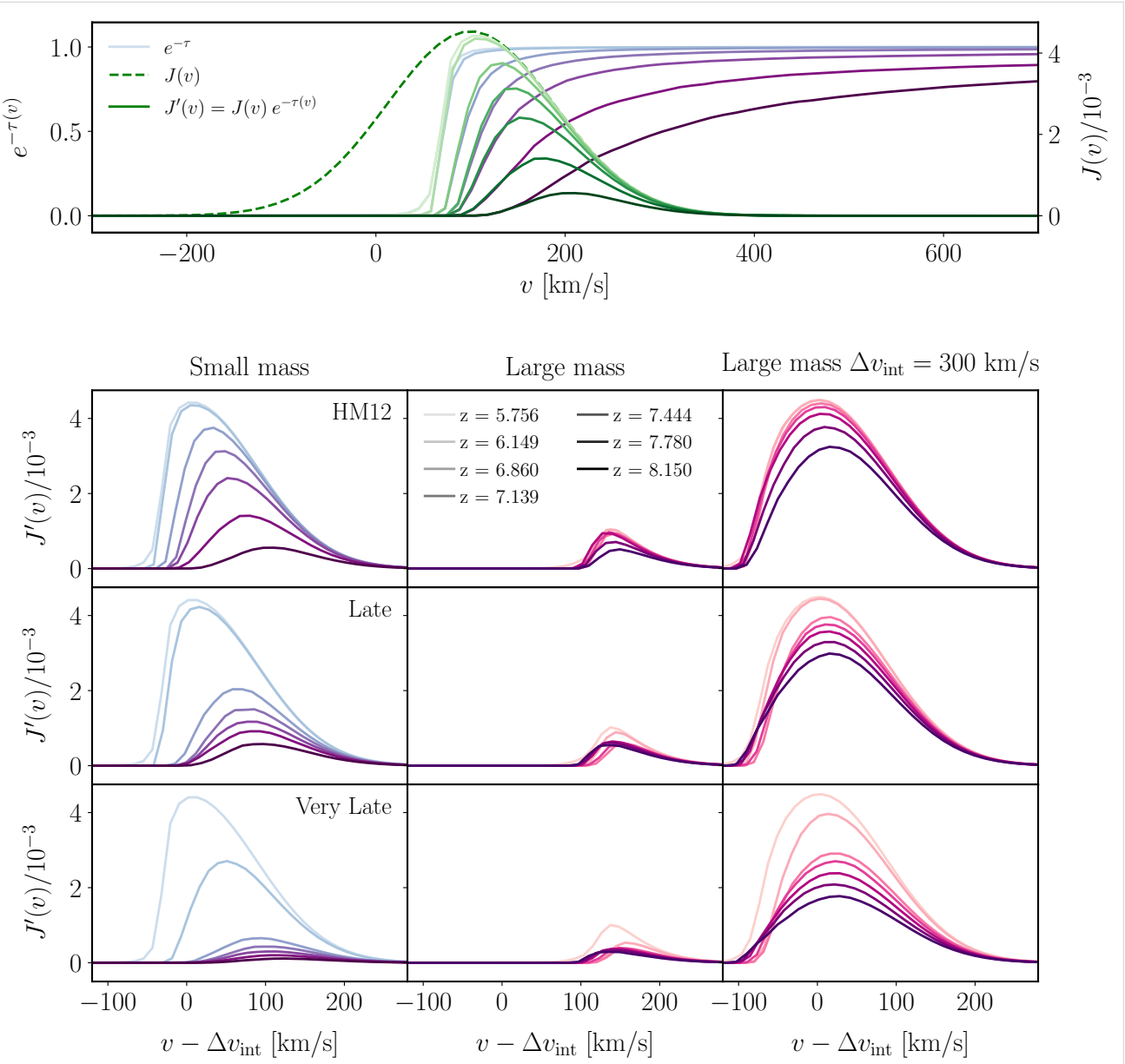


Fig. 3.6 *Top panel*: the (normalised) initial emission profile $J(v)$ (assuming Δv_{int} = 100 km/s, green dashed line) and the median transmission $e^{-\tau(v)}$ (purple solid lines) redward of line-centre (v = 0), between z = 5.756 (light) and z = 8.150 (dark), are shown for the small mass model. The resulting emission profiles (green solid lines) after IGM reprocessing are found as the overlap of these two curves, $J' = e^{-\tau}J$. *Bottom panels*: the resulting median emission profiles for the different mass and reionization history models. The small mass host halo model is shown on the left and the large mass model in the middle panels. The right panels also show the large mass results, but found using a larger intrinsic velocity offset of Δv_{int} = 300 km/s. The reionization histories (HM12, Late and Very Late) are shown from top to bottom.

3.5.1 Evolution of the median transmission

In Figure 3.6 we show the attenuation effect of the IGM on the initial galactic emission profile. The top panel shows the components involved in the transmission fraction calculation: the emission profile in dashed green, the transmission in solid shades of purple (with shade darkening as redshift increases, for the small mass model) and the resulting transmitted emission profile (after IGM reprocessing) in solid shades of green.

The transmission fraction is given by the area under this reprocessed emission profile, as discussed in Section 3.3.1. The lower set of panels show the reprocessed emission profile for 6 of the model combinations: the reionization histories from top to bottom, and the small and large mass models in the left and middle panels respectively. We also show 3 further model combinations in the right hand panels, in which the large mass model is paired with a larger intrinsic velocity offset of 300 km/s than our default 100 km/s. In general the presence of neutral hydrogen gas during the EoR causes the peak of the emission profile to be translated redwards in frequency space, and to be reduced in amplitude. We note that the evolution of the profile is most rapid in the Very Late model. For each reionization history it also occurs more rapidly for the small mass model. The trend for the frequency translation of the profile with redshift is different between the small and large models. The small model profile reddens with increasing redshift. In the large mass model the shift in frequency is less clear. We see that for the same intrinsic emission profile, the resulting profile is more strongly attenuated for the large mass haloes at a given redshift. In the right panels where we have used a larger intrinsic velocity offset ($\Delta v_{\rm int}$ = 300 km/s) we see that instead the large mass halo profiles are less (or equivalently) attenuated compared to the small mass profiles. This demonstrates that the IGM and CGM attenuation of the Lyα luminosity is indeed very sensitive to the intrinsic emission profile. Despite this significant effect seen when comparing at a given redshift, we find that the relative transmission evolution (i.e. normalised to a given reference redshift, as described in Section 3.3.4) is less sensitive to the intrinsic emission profile.

3.5.2 Transmission fraction distribution

We apply the framework from Section 3.3.1 (Eq. 3.3) to explore the difference in the distributions of the transmission fractions for the small and large mass models. In Figure 3.7 we show the (normalised) probability distribution for the transmission fraction at z = 7; we show the small mass (black line) and large mass (red line) models (as well as a large mass model with increased velocity offset of 300 km/s shown with the dotted red line). Considering first the small mass model distribution in black, we see a bimodal distribution with peaks around $T \sim 0$ and $T \sim 0.6$. The $T \sim 0.6$ peak can be understood as those sightlines which start in host haloes sitting in ionised regions, where there isn't sufficient recombined neutral hydrogen (or the neutral gas is not infalling with a high enough velocity) to completely reduce the transmission fraction in the ionised region. The photons emitted in the vicinity of such haloes can redshift beyond the damping wing by the time they reach the

edge of the ionised region, and hence will be transmitted along the sightline. The dominant $T \sim 0$ peak is due to sightlines where photons emitted at the halo position would be absorbed/scattered somewhere along the sightline. This absorption might be due to self-shielded clumps, recombined hydrogen in the ionised regions, or residual neutral hydrogen in the rest of the IGM. Comparing this to the large mass model distribution in solid red, we see instead a single peak around $T \sim 0$, although there is also a small non-zero probability of $T > 0.8$ which wasn't present in the small mass model distribution. Finally the red dotted line shows the same large mass model, but using a larger intrinsic velocity offset of $\Delta v_{\rm int} = 300$ km/s (compared to the default of 100 km/s). This distribution now recovers a second peak at $T \sim 1$. We note that the mean transmission fraction is higher for the sightlines that start on the small mass haloes, unless the larger velocity offset is used for the large mass model haloes.

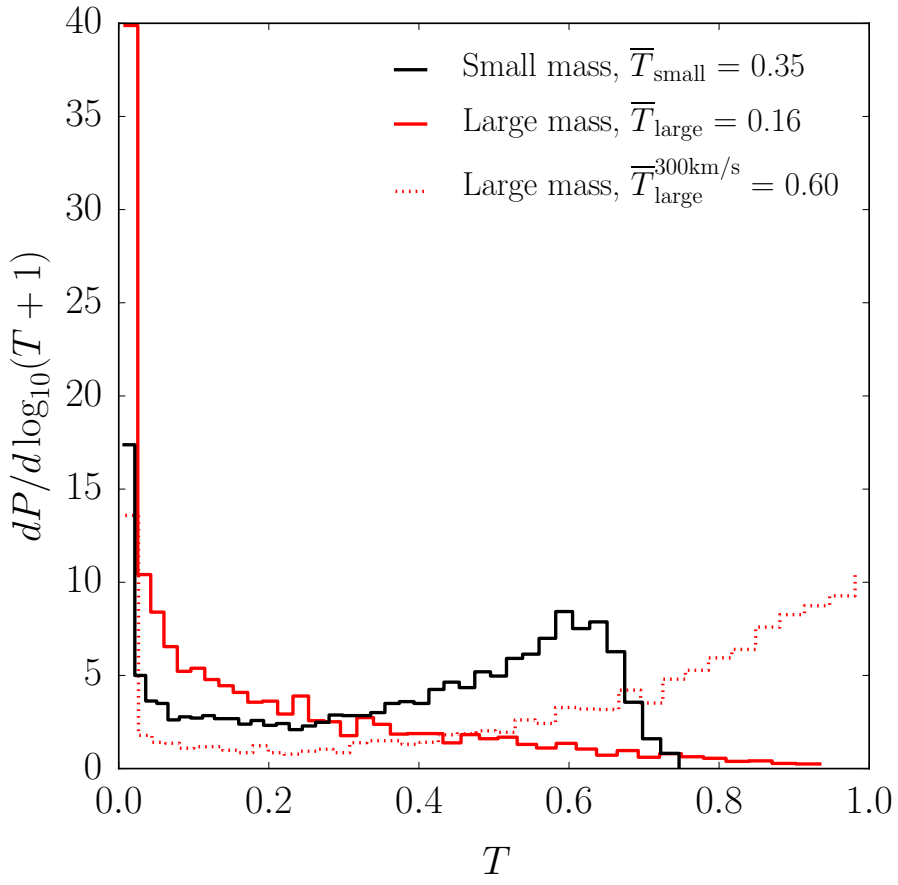


Fig. 3.7 The distribution of transmission fractions for the small (black) and large (red) mass models at $z = 7$, using the fiducial Late reionization history. The red dotted line shows the large mass model distribution if the intrinsic offset is instead $\Delta v_{\rm int} = 300$ km/s.

These distributions may seem counterintuitive, as the more massive haloes should sit in larger ionised regions and hence be more visible on average. This picture however does not take into account the infalling velocities of the neutral gas within ionised bubbles, either recombined or self-shielded,

which are considerably larger for the more massive haloes (as seen in Figure 3.5). This infall towards the halo counteracts the cosmological redshifting of the emitted photons such that they are closer to line centre in the frame of the gas, which leads to greater absorption (unless the intrinsic offset is increased). The $T \sim 1$ peak in the default large mass model (red solid line) is diminished because although these emitters sit in large ionised regions, the self-shielded gas within the ionised region can still strongly attenuate the Lyα emission. However when the intrinsic offset is increased, such that this self-shielded gas becomes more transparent to Lyα radiation, we recover the peak we would expect close to $T \sim 1$.

In this way we see that at a given redshift the presence of neutral CGM gas can lead to an increase in halo-to-halo scatter of the transmission in our mass samples. We note however that the average evolution of the transmission is driven by the neutral IGM. The relative importance of the CGM/IGM absorption in Lyα visibility will be explored further in Section 3.6.2.

3.5.3 Transmission fraction evolution in the small, continuous and large mass models

As discussed in Section 3.3.4, we can quantify the evolution of the transmission fraction by normalising to a reference redshift value (here chosen to be $z = 5.756$), which we call the transmission fraction ratio (TFR). We calculate the mean TFRs at a given redshift for the three mass models, in the three different reionization histories. This can be used to compare how the visibility of LAEs in the different mass models evolves. In Figure 3.8 we plot the TFR evolution of the small mass (cyan) and large mass (magenta) models, with 1 σ scatter shown by the shading. We estimate this scatter by repeatedly sampling the transmission fraction distribution at each redshift, with sample sizes comparable to the observational sample sizes[6]. This results in an increase in scatter with redshift as the sample sizes decrease, reflecting the increase in statistical uncertainty. Beyond redshift $z = 7.3$ the sample size is kept constant, and the scatter starts to decrease as the halo-to-halo variation decreases (because at high redshifts the universe was more homogeneously neutral). In all reionization histories before percolation ($z \gtrsim 6$) we find that the large mass model evolves considerably slower than the small mass model. We see greater scatter in the large mass model; this is likely because the large mass model contains some rare very massive haloes, such that there is a non-negligible difference in environment between the most and least massive haloes within the large mass model. This leads to more halo-to-halo variation in the Lyα transmission along these sightlines, compared to the small mass model (whose mass bin width is smaller).

[6]We note that for their luminosity function samples, Konno et al. (2018) found 1081 LAEs at $z = 5.7$ and 189 at $z = 6.6$.

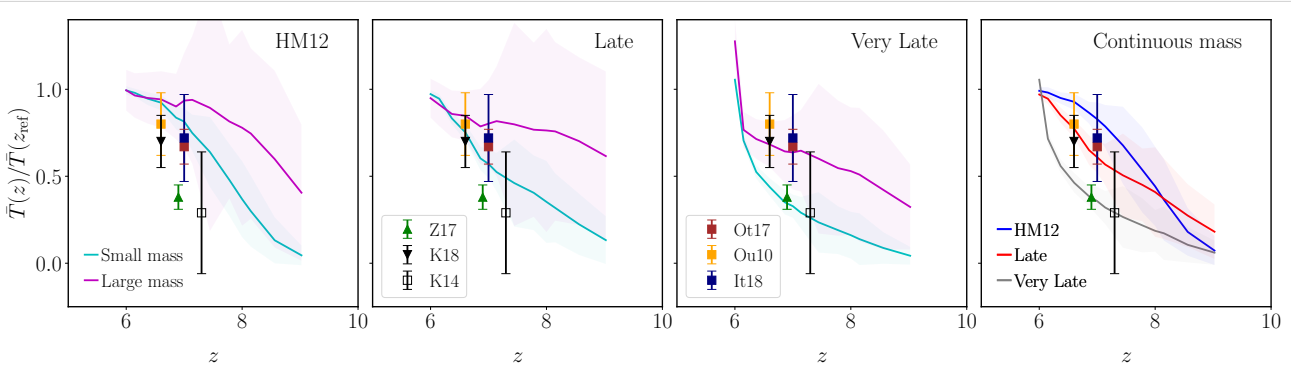


Fig. 3.8 Evolution of the mean transmission fractions (TFRs), normalised to $z_{\rm ref} = 5.756$. *Lefthand three panels*: The large (magenta) and small (cyan) mass models are shown in all panels. The shaded regions show the 68% scatter in the TFR values, found by sampling the distribution with sample sizes comparable to observational LAE samples at each redshift. From left to right we have the different reionization histories: HM12, Late and Very Late. Overplotted is data from Zheng et al. (2017, Z17), Konno et al. (2018, K18), Konno et al. (2014, K14), Ota et al. (2017, Ot17), Itoh et al. (2018, It18) and Ouchi et al. (2010, Ou10) normalised to $z_{\rm ref} = 5.7$; where errors were not quoted in these works we have made a basic estimate. Note these observational data-points were found by considering the luminosity density across the faint ($42 \lesssim \log_{10} L_{{\rm Ly}\alpha} \lesssim 43$) and bright ($\log_{10} L_{{\rm Ly}\alpha} \gtrsim 43$) ends of the luminosity functions. Therefore these observational points are best compared to the continuous model, as shown in the rightmost panel. *Rightmost panel*: The TFR evolution of the continuous mass model. The different reionization history models are shown in blue, red and grey, with the corresponding shading indicating 1σ scatter. This model represents a middle ground between the extreme small and large mass models. Note that in all these panels the emission profile was our default model with $\Delta v_{\rm int} = 100$ km/s.

Overplotted on Figure 3.8 for reference are a selection of observed TFRs reported by Zheng et al. (2017, Z17), Konno et al. (2018, K18), Konno et al. (2014, K14), Ota et al. (2017, Ot17), Itoh et al. (2018, It18) and Ouchi et al. (2010, Ou10), all normalised to $z = 5.7$. Importantly, the TFRs quoted by observers are usually calculated from a full (luminosity spanning) sample of LAEs, i.e. *including* both bright and faint LAEs. As the small and large mass models represent extreme examples of LAEs, the most meaningful comparison with observational data is with our continuous model. Nonetheless we overplot the observational data on all TFR figures, in order to give a reference point for comparison.

In the rightmost panel of Figure 3.8 we show the evolution of the TFRs for the continuous model, with the different reionization histories represented by different colours. We note that the "Late" (in red) and "Very Late" (in grey) reionization histories are the best matches to the observed data, suggesting a reionization history somewhere in between these two bracketing models. The scatter in the continuous model is comparable to the scatter in the small mass model, which in turn is similar to the observational errors.

3.5.4 Differential evolution of the transmission fraction

Alongside the average TFRs reported by observers, some (e.g. Zheng et al., 2017) have also reported that the TFRs for bright LAEs are higher than for faint LAEs. This behaviour is reproduced by our large (representing bright LAEs) and small (representing faint LAEs) models, which show a difference in the TFRs for the same redshifts as seen in Figure 3.8.

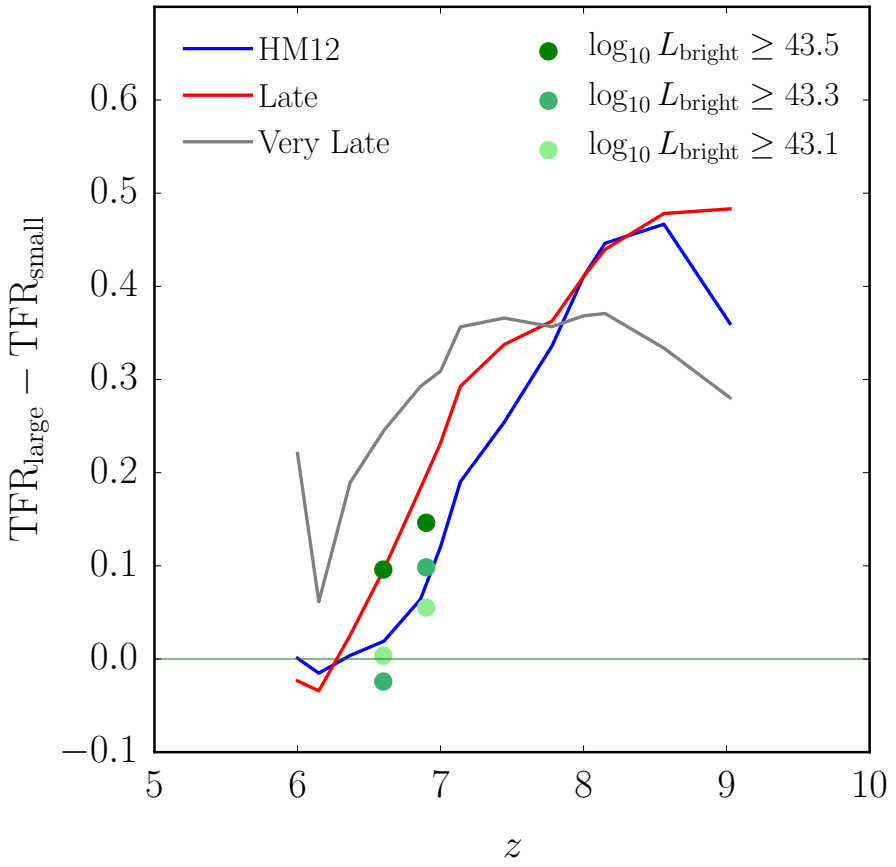


Fig. 3.9 Difference in the evolution of the TFRs (normalised to $z_{\rm ref} = 5.756$) between the large and small mass models. Overplotted in shades of green are observed differences, derived with data from Konno et al. (2014); Zheng et al. (2017); Matthee et al. (2015); Santos et al. (2016) (this data was also normalised to $z_{\rm ref} = 5.7$) Three different brightness thresholds are shown: $\log_{10}(L_{\rm min}^{\rm bright}/\mathrm{erg\ s^{-1}}) = 43.5,\ 43.3,\ 43.1$. Note again that in all these models the emission profile was our default model with $\Delta v_{\rm int} = 100$ km/s.

To compare this more explicitly, we use reported observational data from Konno et al. (2014); Zheng et al. (2017); Matthee et al. (2015); Santos et al. (2016) to reconstruct the TFRs for the bright and faint LAEs separately. In this way we want to establish trends and obtain a lower limit on the bracketing values for the TFRs, and so do not perform a detailed re-analysis of the data. We take

values of $\Phi(L) = \mathrm{d}n/\mathrm{d}\log_{10} L$ as quoted in the original works. From these we calculate,

$$\rho(z) = \int_{L_{\rm min}}^{L_{\rm max}} L\,\Phi(L)\,d\log_{10} L. \tag{3.16}$$

In order to perform this integral we apply a trapezoidal algorithm on the published data points; we do not fit a Schechter form. The data is heterogeneous in terms of the luminosity ranges observed, so we impose limits, $L_{{\rm Ly}\alpha} \in [42.5, 43.7]$ erg/s, and use linear interpolation and extrapolation to evaluate each of the datasets in the same luminosity bins along this range. There is obviously freedom in the choice of the "bright" threshold; we test values around $\log_{10} L_{\rm min}^{\rm bright} = 43$, bracketing $\log_{10} L_{\rm min}^{\rm bright} = 43.4$ as used in Zheng et al. (2017). The threshold used for the calculated values plotted in Figure 3.9 are $\log_{10}(L_{\rm min}^{\rm bright}/{\rm erg\ s^{-1}}) = 43.1,\ 43.3,\ 43.5$. We then calculate the TFRs using the expression in Eq. (3.12), with the UV data from Finkelstein et al. (2015)[7], assuming $\kappa(z = 5.7)/\kappa(z = 6.6) = 1$ and $f_{{\rm esc,Ly}\alpha}(z = 5.7)/f_{{\rm esc,Ly}\alpha}(6.6) = 1$.

The green markers in Figure 3.9 show the (re-calculated) difference in the TFRs for the bright and faint LAEs, compared to the simulated differences between the large and small mass models (lines coloured by reionization history). The differential TFR evolution depends on the chosen reionization history, but the shape of this evolution is similar across the models. We will discuss this further in section 3.6.1. Note that changing the bright threshold in the observed data alters the amplitude of the difference, and the slope across redshifts. [8]

3.5.5 Evolution of the Lyα fraction of LBGs

Finally, we also consider the independent observational measurement of the Lyα fraction of LBGs, $X_{{\rm Ly}\alpha}$, to see if our large and small mass models can be used to reproduce the UV bright and UV faint evolution. We calculate this evolution as described in Section 3.3.5.

In Figure 3.10 we compare the evolution of $X_{{\rm Ly}\alpha}$ predicted by our simulations with the observed data, for the thresholds of REW > 25 Å and REW > 55 Å. Our models are again reasonably consistent with the data; the largest discrepancy is found for the steep drop in the REW > 25 Å UV faint data of Stark et al. (2011) which only our 'Very Late' model is able to reproduce. We note that the use of the large mass model for the UV bright data accounts for the slower decline in this sample, whilst the faster evolution of our small mass model is a good fit for the UV faint sample. Apart from the left panel (REW > 55 Å UV faint), the comparison with observational data does not exclude any of our reionization history models.

[7] We also calculated using UV data of Bouwens et al. (2015), but the bright/faint trend persists regardless of this change.

[8] We note that the simulated TFR difference is sensitive to the chosen mass bins, and hence without better constraints on host halo masses the observed TFR differences cannot be used to constrain the most likely reionization history. Our large and small mass models are nevertheless useful for demonstrating that a difference does indeed occur.

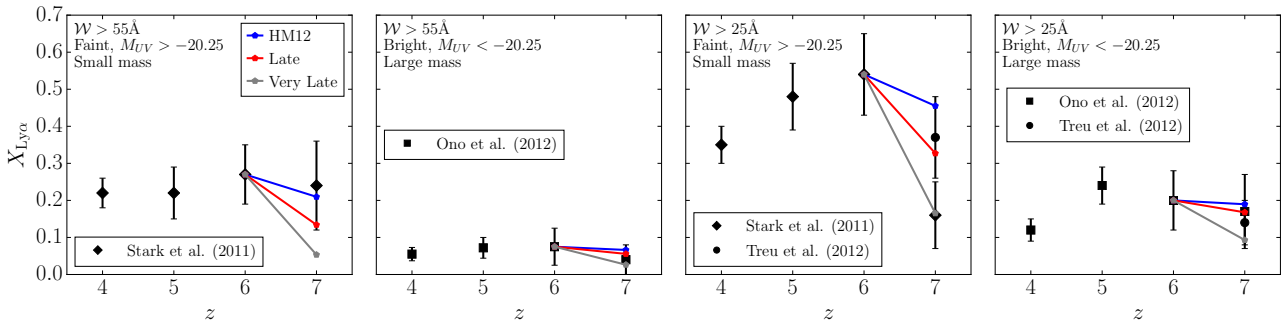


Fig. 3.10 The fraction of galaxies emitting Lyα above a given threshold equivalent width REW (EW), as a function of redshift. These distributions were calculated using the small mass model (first and third panels, UV faint) and large mass model (second and fourth panels, UV bright). The different reionization models are shown as blue (HM12), red (Late) and gray (Very Late) lines. For comparison the observed *UV faint* ($M_{\rm UV} > -20.25$) data from Stark et al. (2011, diamonds), Treu et al. (2012, circles) and *UV bright* ($M_{\rm UV} < -20.25$) data from Ono et al. (2012, squares), Treu et al. (2012, circles) are also shown.

Comparing with SZM17 (Sadoun et al., 2017), we see similar predictions to their infall model, despite having a more modest evolution of $\Gamma_{\rm HI}$[9] in our simulations.

3.6 Discussion

In Section 3.5.4 we have shown that our simulations predict a difference in the evolution of the visibility of LAEs hosted in different mass haloes. If we assume that indeed brighter LAEs are found in more massive host haloes, then this can explain the different evolution of bright and faint high-redshift LAEs. We now discuss possible physical mechanisms for this difference in our simulations. We caution, however, that some of the observed difference could also be due to observational selection effects.

3.6.1 Differential evolution of large and small mass models

Neglecting intrinsic galaxy evolution, we explore here two different aspects of the IGM and CGM attenuation that might cause the different evolution of the bright and faint LAE populations.

1. The most intuitive mechanism is perhaps the different (large-scale) environments of ionised bubbles in which LAEs might reside (for example discussed in Section 4.1 of the study by Ota et al., 2017). More massive haloes are likely to reside in larger ionised regions compared to less massive haloes. In particular we might also expect that (depending on the reionization history) more massive haloes will be surrounded by ionised regions earlier, after which their visibility

[9] Their model considers a change between $z = 7 \longrightarrow 6$ in the photoionisation rate of $\Gamma_{\rm HI} = 10^{-14} \longrightarrow 10^{-13}$ s^{-1}; the minimum $\Gamma_{\rm HI}$ in our models between $6 < z < 8$ does not fall below $10^{-13.2}$ s^{-1}.

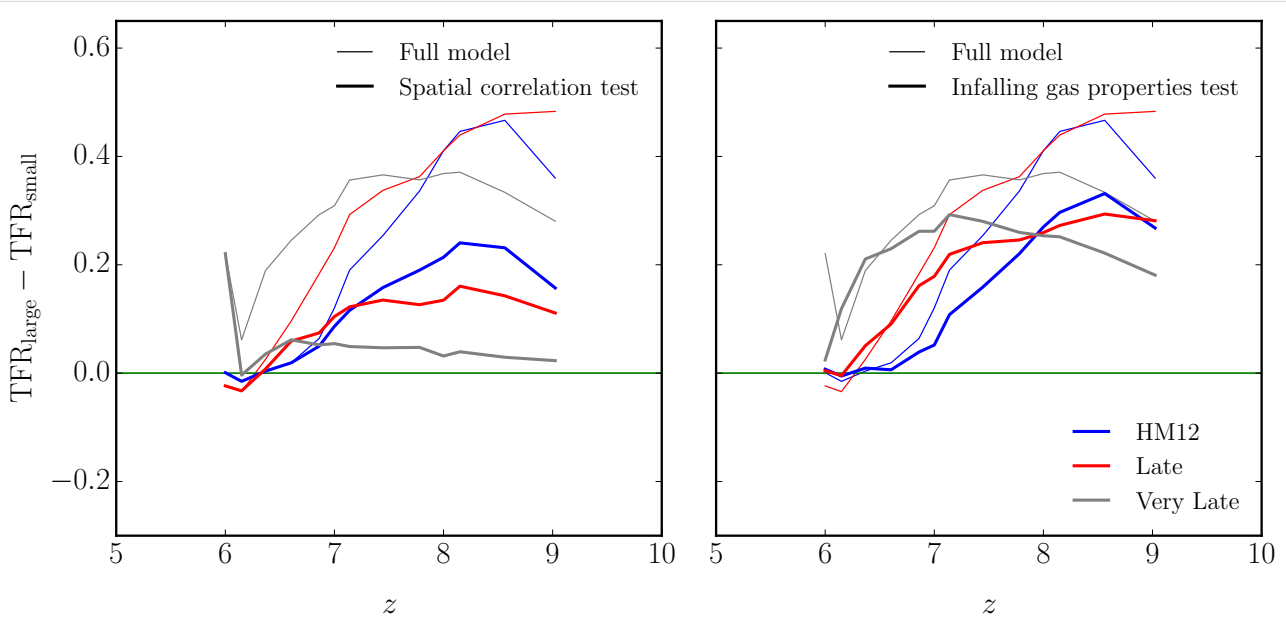


Fig. 3.11 Evolution of the ratio of transmission fractions, testing the effect of the CGM/IGM on the differential visibility of large and small host haloes. The difference in TFR between the large and small mass models is plotted for: the full models (thin lines), the test cases (thick lines). In the left panel we show the "Spatial correlation test", whilst in the right panel we show the "Infalling gas properties test". The different reionization history models are shown in blue (HM12), red (Late) and gray (Very Late).

will not evolve dramatically; in comparison the less massive haloes will enter overlapping ionised regions around the more massive haloes at later times.

2. A second, more subtle mechanism is due to the different dynamical properties of neutral hydrogen in the CGM. In Figure 3.5 we showed the evolution of the infall velocity of gas around haloes of different masses. We might expect both the gas close to the halo (which includes self-shielded or recombined neutral hydrogen within the ionised region) and the residual neutral gas in the not-yet-ionised IGM around the halo to absorb differently depending on the host halo mass.

In order to explore the contributions of these two mechanisms in our models, we perform the following two tests.

- *Spatial correlations with ionised bubbles*: we displace the ionisation field along one direction of the simulation by a distance d = 80 cMpc/h, half the simulation box length. This will break the correlation between the location of haloes and ionised regions. If the higher TFRs of the more massive host haloes are caused by their position in larger ionised bubbles, then this test should result in the TFRs of the small mass and large mass models converging.

- *CGM peculiar velocities and temperature*: we recalculate the optical depths along the extracted sightlines, but neglect both the peculiar velocities and the temperature variation of the CGM.

Note, however, that the two mechanisms are coupled and that the two tests are therefore not independent of each other. In particular in the first test, by displacing the ionisation field we will also be removing some of the correlation between the in-falling velocities of neutral gas and the haloes. We should therefore not expect the two tests to quantify how much each of the mechanisms is contributing to the difference in TFR evolution, but they should nevertheless show whether these two mechanisms are indeed having an effect.

In Figure 3.11 we plot the difference between the TFRs for the large and small mass models as a function of redshift. The left panel shows the spatial correlation test. The difference drops close to zero for all reionization histories suggesting that indeed the difference of the visibility of smaller and large mass haloes decreases significantly if there is no correlation of their location with that of ionised bubbles. We note that the effect of the correlation depends on the reionization history. The strongest effect of removing the correlation is seen in the Very Late model and it is weakest for the HM12 model.

This can be understood by considering the rate with which smaller haloes enter the large-scale overlapping ionised regions. The overlapping ionised regions initially develop around the largest haloes which provide a bigger fraction of the total ionising photon budget, and hence these haloes remain in ionised regions out to higher redshifts. Smaller haloes enter into these ionised regions later, when the ionisation fronts around the larger haloes percolate and expand into the ionisation fronts around these smaller haloes. How quickly the smaller haloes enter the ionised regions depends strongly on the reionization history, both on $\langle x_{\rm HII}(z)\rangle_{\rm M}$ and $\mathrm{d}\langle x_{\rm HII}(z)\rangle_{\rm M}/\mathrm{d}z$. In the HM12 model reionization ends early such that around $z \sim 7$ both the small and large halo positions are strongly correlated with the ionised regions. This means the difference in visibility of these haloes is mostly not determined by the sizes of the ionised bubble. In the Very Late model, in which reionization ends later and more rapidly, there is a much larger difference in the bubble sizes surrounding the small and large haloes at $z \sim 7$.

In the right panel Figure of 3.11 we show the effect of neglecting gas peculiar velocities and temperatures of the CGM surrounding the host haloes. Neglecting these gas properties, there will be less absorption in neutral hydrogen around the host halo (both within reionized regions and also in the residually neutral IGM). For this test there is also a dependence on the reionization history, however for all our reionization models the effect is less significant than that of removing the spatial correlation with the ionised bubbles. The influence of the infalling gas properties increases with redshift. For example the fractional difference from the full calculation in the Late model is ~ 0.3 at $z = 8$, but rises to ~ 0.4 by $z = 9$.

In summary, the results of our two tests suggest that, dependent on the reionization history:

- the positions of more massive haloes in larger ionised regions can make a significant contribution to the differential visibility of the large and small mass models.
- the infalling gas properties of neutral IGM gas also play an, albeit smaller, role in the increased visibility of the large mass model.

We also note that the largest difference in visibility occurs for the Late reionization history.

3.6.2 The effect of self-shielding and the dominant scales on which IGM attenuation occurs

In SZM17 (and also in earlier work such as Kakiichi et al., 2016; Dijkstra et al., 2007), the role that the infalling CGM gas plays in the Lyα attenuation was explored. We have seen in Figure 3.3, that the self-shielded gas in the CGM can indeed attenuate Lyα alongside the more distant neutral gas in the not yet ionised regions of the IGM. The strength of the attenuation depends on the amount of self-shielded gas present, and hence also on the local photoionisation rate. As the global neutral fraction of the large-scale IGM is also coupled to the photoionisation rate, we note that these two attenuating components are also coupled. Within our models, the strength of the attenuation due to the self-shielded gas in the CGM will depend on the assumed self-shielding prescription, the amount of gas that is excluded from within the host halo, and the intrinsic velocity offset of the Lyα emission profile. In this subsection we aim to explore the interplay between this inner CGM self-shielded gas and the external (residual) neutral IGM gas, to try to quantify the strength of the roles that they play in attenuating Lyα from high redshift galaxies.

In Figure 3.12 we show how the transmission fraction at $z = 6$ depends on the background photoionisation rate. In all of our reionization histories at this redshift, the IGM is ionised ($\langle x_{\rm HII} \rangle_{\rm M} = 1$), and so only the self-shielded/recombined CGM gas can play a role. In order to quantify how strong the attenuation can be from this gas, we normalise the transmission fraction to the value for $\Gamma_{\rm HI} = 10^{-12}$ s^{-1}. In each of the three panels of Figure 3.12 we then test the effects of our assumptions: on the left the self-shielding prescription, in the middle the exclusion regime and on the right the emission profile offset. In all panels we see that decreasing the background photoionisation rate (and therefore increasing the amount of self-shielded gas) increases the attenuation of Lyα. We note that in our fiducial reionization history, however, the background photoionisation rate doesn't fall lower than $\Gamma_{\rm HI} \sim 10^{-13.2}$ s^{-1}.

In this work we have employed the self-shielding prescription suggested by Chardin et al. (2018) (labelled SS-Ch). Other works have used different prescriptions for self-shielding, which can lead to more neutral gas and thus a stronger attenuation of Lyα emission. In the left panel of Figure 3.12 we compare our fiducial prescription with: (i) the case of no self-shielding, and (ii) with a prescription based on Rahmati et al. (2013) (the default choice of CPBH15, labelled SS-R). As expected we see that the stronger the self-shielding the more attenuation can come from this CGM gas. However even in the case of no self-shielding, where the amount of neutral gas is given only by recombinations in photoionisation equilibrium, we see that if the photoionisation drops sufficiently then the transmission fraction can be reduced. In the middle panel we see the effect of excluding different amounts of the CGM gas. Importantly we note that for our default SS-Ch self-shielding prescription and photoionisation rates larger than $\Gamma_{\rm HI} \sim 10^{-13}$ s^{-1} (that are suggested by full radiative transfer simulations of reionization), the attenuation is not very sensitive to the size of the exclusion

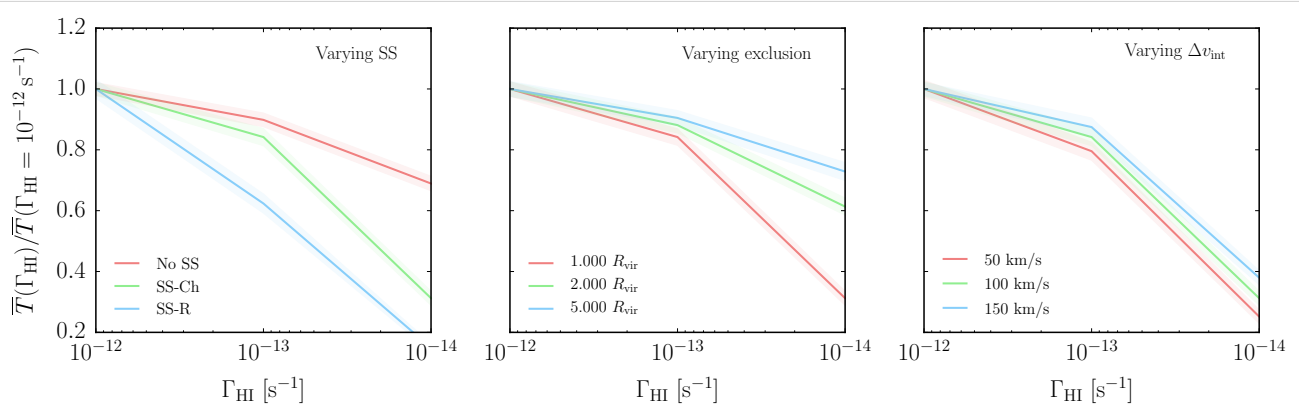


Fig. 3.12 The photoionisation rate dependence of the transmission fraction at $z = 6$, normalised to the value when $\Gamma_{\rm HI} = 10^{-12}$ s^{-1}. *Left*: The effect of varying the self-shielding prescription. *Middle*: The effect of varying the amount of gas that is excluded around the halo position. *Right*: The effect of varying the intrinsic velocity offset of the emission profile. Shading indicates 68% scatter around the mean, calculated as in Figure 3.8.

region. In the righthand panel we show the effect of changing the intrinsic velocity offset of the emission profile. We see that of the three assumptions tested in this figure, the results are least sensitive to this choice.

Note that, for the photoionisation rates in our reionization histories, the self-shielded CGM gas alone can attenuate the Lyα signal by as much as ~ 30% for the Rahmati et al. (2013) self-shielding. For this model the dependence on the size of the exclusion region is therefore also stronger than our default self-shielding model.

In order to explore this further, we also show the effect of changing our assumptions for the full TFR evolution in Figure 3.13, using the continuous mass model and the Late reionization history. This therefore includes the contributions of both the CGM and the IGM. As in the previous figure, on the left panel we show the effect of the self-shielding prescription, in the middle we show the effect of the exclusion region, and in the right panel we show the effect of the velocity offset. In the left panel we also include the prescription used in Bolton & Haehnelt (2013) (labelled SS), which assumes a sharp threshold for self-shielding at the Jeans scale. The results found without self-shielding can be considered as the attenuation due to the residual neutral IGM alone. We see that the neutral IGM is the dominant component in determining the average redshift dependence of the attenuation. However the self-shielded gas can also play an important role, depending on the self-shielding prescription (SS resulting in the most self-shielding, and SS-Ch the least). In the central panel we see the effect of excluding different amounts of the CGM gas. For exclusion regions > $2R_{\rm vir}$, the TFR depends very weakly on the exact choice of exclusion radius and values close to those in the No-SS case (shown in the left panel) are found. Finally we see in the right hand panel that varying the intrinsic velocity offset does not alter the TFR evolution very much. Although the transmission at a given redshift might be sensitive to these changes, the normalisation of the TFR removes part of this sensitivity (so long as the velocity offset is independent of redshift).

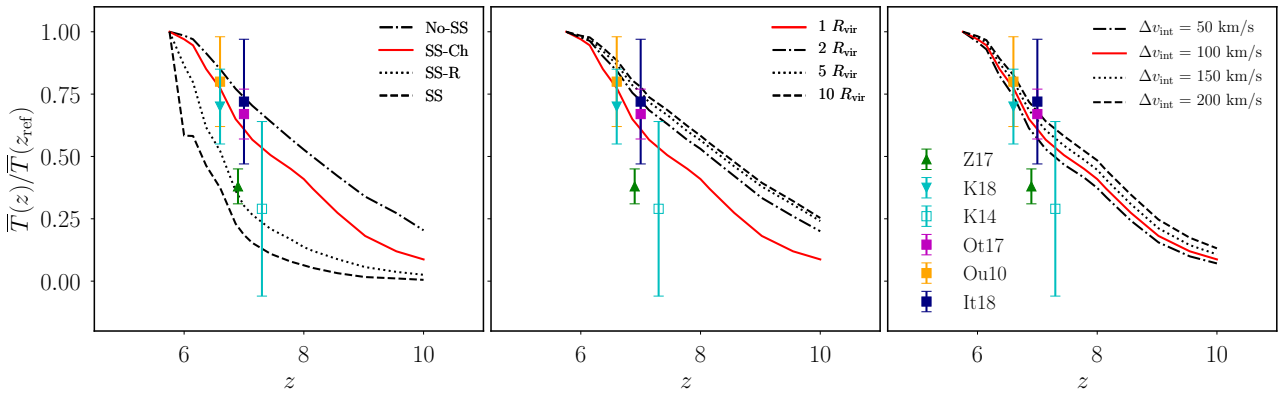


Fig. 3.13 Testing the effects of different assumptions on the TFR evolution. *Left*: The effect of a range of self-shielding prescriptions, using the Late reionization history and the continuous model. The fiducial prescription used in this work is the SS-Ch model based on Chardin et al. (2018), shown in red. *Middle*: The effect of excluding gas around the halo, within a number of virial radii. Our fiducial exclusion is 1 $R_{\rm vir}$, shown in red. We note that beyond an exclusion of 2 virial radii, the evolution is very insensitive to the exact choice of exclusion radius, however there is still a strong attenuation. This remaining attenuation is due to the large-scale neutral IGM, as previously modelled. *Right*: The effect of changing the intrinsic velocity offset of the emission profile. Our fiducial offset is 100 km/s, shown in red. Overplotted are observed TFR values, as in Figure 3.8.

3.6.3 Observational selection effects

Throughout this work we have relied on the basic assumption that there is a positive correlation between the host halo mass and a galaxy's (rest-frame) Lyα luminosity. Lyα photons are created in a galaxy's ISM by reprocessing the ionising photons emitted from the stellar component. The Lyα luminosity depends on the star formation rate (SFR), which in turn depends on the host halo mass, M_h, (Zheng et al., 2010). Given the often bursty nature of star formation it is nonetheless not obvious that the brightest LAEs are hosted in the most massive haloes.

In the first instance we have calculated the TFR evolution, and compared to narrowband Lyα selected galaxies (such as in Ouchi et al., 2010). We split the samples into *bright* and *faint* based on the observed Lyα luminosity. For this selection method a galaxy might be categorised as a bright LAE but might not necessarily be hosted by a more massive host halo. This is because the flux in the Lyα narrowband filter is compared with a (sometimes overlapping) broadband filter; the galaxy may appear bright with this selection method because there is more flux in the narrowband than in the UV continuum. This therefore includes cases where the UV continuum is faint, and hence the galaxy may be less massive.

We have also calculated the evolution of $X_{{\rm Ly}\alpha}$, and compared to dropout selected galaxies with spectroscopically confirmed Lyα equivalent widths (REW) above a given threshold (such as in Ono et al., 2012). These galaxies are first selected using the Lyman break technique, and divided into UV-bright and UV-faint, based on bolometric UV luminosity. This UV luminosity correlates with

stellar mass, and hence the UV-brighter objects will be hosted in larger mass haloes. The secondary Lyα equivalent width selection does not change this measurement, so in this case the brighter LAEs will almost certainly correspond to more massive haloes. Dayal & Ferrara (2012) have suggested that indeed the $z > 6$ LAEs form a luminous subset of LBGs.

The applicability of our different mass models, and in particular the mapping from these models to the different populations of LAEs (divided by brightness), is therefore dependent on the way the population is selected. The TFRs we have calculated using the continuous mass model are probably the most realistic. For the $X_{\mathrm{Ly}\alpha}$ evolution however, our application of the different mass models to the different UV brightness samples is probably better justified.

3.7 Conclusions

We have updated the modelling of the rapid evolution of Lyα emitters by CPBH15 (Choudhury et al., 2015) using the high-dynamic range Sherwood simulations as a basis for our analytical model for the growth of ionised regions. We have in particular assessed the effect of host halo mass on LAE visibility just before the percolation of HII regions occurs at $z \sim 6$. Our main results can be summarised as follows:

- Our simulations naturally reproduce the observed strong difference in the evolution of the visibility of bright and faint LAEs at $z \geq 6$ if we assume that bright LAEs are placed in the most massive haloes in the simulations with similar space densities as observed for bright LAEs.

- The less rapid evolution of the visibility of bright LAEs in our simulations at $z > 6$ is only partially due to their strong spatial correlation with the first regions to be reionized, an explanation that has been invoked by other authors. In our simulations we find an additional contribution: the different gas peculiar/infall velocities and peak temperatures in the environment of massive haloes contribute to the differential evolution of bright and faint LAEs. The relative contribution of the evolution of peculiar/infall velocities and the spatial correlations with ionised regions on the visibility of LAEs thereby depends strongly on the assumed reionization history.

- It is the faint emitters that more closely trace the evolution of the volume-filling fraction of ionised regions, since the gas in their local environments is not rapidly evolving (as it is for the bright emitters). We thus recommend that studies of the reionization history continue to focus on the fainter LAEs.

- In our simulations the infalling gas in the outskirts of the halo (just outside the virial radius) has a strong effect on the visibility of the LAE it is hosting. This is in agreement with the suggestion by Sadoun et al. (2017) that before percolation the infalling gas in the outskirts of LAE host haloes in already ionised regions is still sufficiently neutral to cause a rapid evolution of LAE visibility at $6 < z < 7$. In our simulations the photoionisation rate in ionised regions

is higher than was modelled in that work, but the self-shielding is still sufficient to strongly attenuate the Lyα emission from the galaxy. In particular we find that this effect is stronger in the more massive haloes. This means that for observations of UV bright galaxies living in such massive hosts, deriving constraints on the volume-filling neutral fraction of the IGM involves more complicated modelling of such self-shielding than for UV faint LAEs living in less massive haloes. This reinforces our recommendation that future observational studies focus on UV faint LAEs for constraining reionization. Alternatively, selecting LAEs based on intrinsic velocity offset could sample those galaxies whose emission is least attenuated by the self-shielded gas of the CGM.

- Overall our updated modelling with the higher dynamic range Sherwood simulation gives similar results to CPBH15, albeit with some notable differences:

    1. We confirm that the 'Late' and 'Very Late' reionization histories favoured in CPBH15, which also match Lyα forest data, are a good match to the observed rapid evolution of faint Lyα emitters. Note, however, that unlike CPBH15 we can obtain this agreement without invoking an evolution of the redshift of the intrinsic Lyα emission relative to systemic. This is possibly due to the more consistent treatment of peculiar velocities in our simulations made possible by dropping the hybrid approach of CPBH15 (who combined a rather small box-size hydrodynamical simulation with a large box-size dark matter simulation). We further confirm that the evolution of the ionising emissivity in the popular HM12 UV background model corresponds to a decrease of the volume factor of ionised regions at $z > 6$ that is too slow to explain the rapid disappearance of faint LAEs.
    2. As in CPBH15, in our updated simulations the rapid decrease of the visibility of faint Lyα emitters is mainly due to the rapid evolution of the volume-filling fraction of ionised regions in our models. In our fiducial updated model we have used the self-shielding prescription suggested by Chardin et al. (2018) who have explicitly modelled the self-shielding in ionised regions before the full percolation of ionised regions with full radiative transfer simulations. Note in particular that with this prescription the effect of self-shielding is significantly weaker than with the widely used Rahmati et al. (2013) model. If self-shielding is indeed as weak as suggested by the Chardin et al. (2018) simulations, then reproducing the rapid evolution of faint Lyα emitters at $z > 6$ may require a reionization history where reionization occurs as late as in our "Very Late" model.

The rapid disappearance of faint Lyα emitters arguably provides the strongest constraints to date on the reionization history of hydrogen at $z > 6$, and our simulations confirm that their rapid disappearance is strong evidence for a rather late reionization.

Chapter 4

Modelling the Observed Luminosity Function and Clustering Evolution of Lyman-α Emitters: Growing Evidence for Late Reionization

In this chapter we model the high redshift ($z > 5$) Lyman-α emitting (LAE) galaxy population using the empirical rest-frame equivalent width distribution. We calibrate to the observed luminosity function and angular correlation function at $z = 5.7$ as measured by the SILVERRUSH survey. This allows us to populate the high-dynamic-range Sherwood simulation suite with LAEs, and to calculate the transmission of their Lyα emission through the intergalactic medium (IGM). We use this simulated population to explore the effect of the IGM on high-redshift observations of LAEs, and make predictions for the narrowband filter redshifts at $z = 6.6$, 7.0 and 7.3. Comparing our model with existing observations, we find a late reionization is suggested, consistent with the recent low optical depth derived from the cosmic microwave background (CMB) by the Planck Collaboration and the opacity fluctuations in the Lyα forest. We also explore the role of the circumgalactic medium (CGM) and the larger volume of gas which is infalling onto the host halo versus the IGM in attenuating the Lyα signal, finding that a significant fraction of the attenuation is due to the CGM and infalling gas, which increases towards the end of reionization, albeit with a large scatter across the mock LAE population.

Declaration

The work presented in this chapter was published as a paper titled "*Modelling the observed luminosity function and clustering evolution of Lyα emitters: growing evidence for late reionization*", in Monthly Notices of the Royal Astronomical Society, Volume 485, Issue 1, p.1350-1366. This work was done in collaboration with Martin G. Haehnelt and Girish Kulkarni. As in Chapter 3, the underlying simulations used for the modelling were run by a prior collaboration (Bolton et al., 2017); otherwise the modelling and analysis were entirely performed by this author, with guidance provided by the collaborators.

4.1 Background

Observational studies of Lyman-α emitting galaxies (LAEs) at high redshifts, $z \geq 5$, have now amassed a considerable population of objects that can be used to learn much about the reionization era and galaxy evolution. These include widefield narrowband surveys such as Itoh et al. (2018); Konno et al. (2018); Zheng et al. (2017); Bagley et al. (2017); Ota et al. (2017); Santos et al. (2016); Matthee et al. (2015); Konno et al. (2014); Kashikawa et al. (2011); Ouchi et al. (2010, 2008), probing redshifts z = 5.7, 6.6, 7.0 and 7.3 using filters on instruments such as the SuprimeCam and HyperSuprimeCam of the Subaru telescope (Miyazaki et al., 2002, 2012). Spectroscopic studies such as Mainali et al. (2018); Matthee et al. (2017); Diener et al. (2017); Shibuya et al. (2014); Pentericci et al. (2014); Ono et al. (2012); Schenker et al. (2012); Stark et al. (2011, 2010); Kashikawa et al. (2006) have allowed detailed confirmation of some of the most interesting objects, including CR7 (Sobral et al., 2015, 2017) and COLA1 (Hu et al., 2016; Matthee et al., 2018), as well as a better understanding of how LAEs fit into a bigger picture of high-z galaxies.

LAEs have been established as a key tool for understanding the progress of reionization at $z \geq 6$. Given the resonant scattering of Lyα by neutral hydrogen in the intergalactic medium (IGM) and circumgalactic medium (CGM) (Gunn & Peterson, 1965; Madau & Rees, 2000), the visibility of LAEs should decrease as observations probe further into the reionization era (Dijkstra et al., 2007). This attenuation can be seen in the redshift evolution of both the luminosity function (e.g. Konno et al., 2018) and the clustering signal (e.g. Ouchi et al., 2018). Studies at lower redshifts are also important for understanding the ionising photon budget, for example Nakajima et al. (2018) used LAEs to understand ionising photon escape fractions, to determine if LAEs could have played a significant role in (re)ionising their surrounding IGM. Green Pea galaxies, low-redshift analogs of LAEs, present a further avenue for understanding the properties of LAEs (Yang et al., 2016).

There has been extensive theoretical modelling of LAEs, considering both the 'intrinsic' Lyα emission properties of galaxies as a result of radiative transfer within the halo (such as Gronke & Dijkstra, 2016; Dijkstra, 2014; Zheng et al., 2010; Dijkstra & Westra, 2010; Laursen et al., 2009; Partridge & Peebles, 1967) as well as the effect of further attenuation by resonant scattering with neutral gas in the CGM and IGM (such as Laursen et al., 2019; Mason et al., 2018b; Inoue et al., 2018; Weinberger et al., 2018; Sadoun et al., 2017; Kakiichi et al., 2016; Hutter et al., 2015; Mesinger et al., 2015; Choudhury et al., 2015; Hutter et al., 2014; Jensen et al., 2014, 2013; Bolton & Haehnelt, 2013; Laursen et al., 2011; Dijkstra et al., 2011; Dayal et al., 2009). In particular these two regimes are often modelled differently: the escape of Lyα photons from within a galaxy is frequently treated using full radiative transfer post-processing of high resolution hydrodynamic simulations (e.g. Verhamme et al., 2006), whilst the scattering in the IGM can be well approximated using $e^{-\tau}$ models (e.g. Laursen et al., 2011). On top of this it is important to model how LAEs form part of the wider galaxy population; Dijkstra & Wyithe (2012) for example modelled LAEs as a subset of the Lyman break galaxy (LBG) population with a M_{UV}-dependent distribution of Lyα equivalent widths. Given a mapping between host halo mass and galaxy UV luminosity, it is possible to employ such a model to fit the UV and

Lyα luminosity function evolution, whilst also being consistent with observed equivalent width distributions. There is however some degeneracy in the mapping from mass to M_{UV}, which can be broken by considering the spatial clustering of LAEs (Lee et al., 2009). It is possible to fit all these observational constraints if a duty cycle is employed (Trenti et al., 2010).

In Chapter 3 we employed large, high-dynamic-range hydrodynamical simulations of the IGM in combination with a semi-analytic reionization model in order to calculate the transmission of Lyα from LAE host haloes. The effects of the CGM, self-shielded neutral gas, and host halo mass dependence on the transmission were explored. In this chapter we now extend those transmission models to include an empirically-constrained model for the population of LAEs. Our modelling updates previous work and employs large scale simulations of the IGM to test the effects of reionization at observable redshifts, and to predict the LAE luminosity function, angular correlation function and rest-frame equivalent width probability distribution simultaneously. The underlying simulations have been used to model the 21 cm signal during reionization (Kulkarni et al., 2016) as well as opacity fluctuations after reionization (Kulkarni et al., 2019a), such that we will be able to perform like-for-like comparisons with these other reionization observables in future work. This allows us to make predictions for the evolution of the Lyα luminosity function and clustering signal, which we compare to available observational data.

This chapter is structured as follows: in Section 4.2 we outline how we model the LAE populations using a suite of cosmological hydrodynamic simulations, then in Section 4.3 we present a comparison of the predictions of our models with current observed data. In section 4.4 we discuss the assumptions of our modelling, before concluding in Section 4.5.

4.2 Methods

4.2.1 Numerical simulations of the IGM

In order to model the IGM gas properties we employ the high-dynamic-range Sherwood simulation suite of cosmological hydrodynamic simulations (Bolton et al., 2017), taking the same approach (detailed in Section 2.2) as in Chapter 3.

In this chapter we utilise a simulation box of side length $L = 320$ cMpc/h and particle number $N = 2 \times 2048^3$, so as to be large enough to capture the clustering signal at large scales. The gas properties were smoothed onto a uniform grid with cell size $L_{\rm cell} = 156.25$ ckpc/h, using the SPH kernel. We employ the calibrated excursion set framework described in Section 2.2.2 in order to post-process these simulations with our desired reionization histories. We point interested readers to the additional material in Section 4.6.1, where we compare this excursion set scheme to a full radiative transfer calculation, and find that our results are largely insensitive to the use of this approximation.

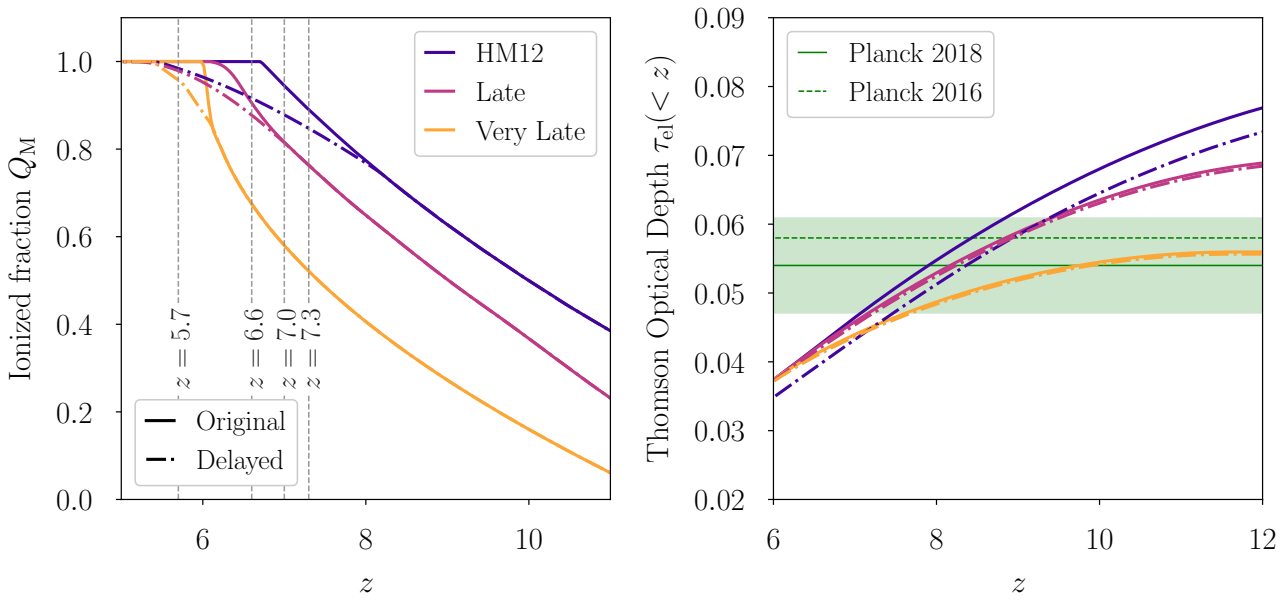


Fig. 4.1 Left: Mass-averaged global ionised fraction for the reionization histories tested in this work, labelled HM12 (blue) Late (purple), Very Late (orange). The solid lines show the original models, whilst the dash-dotted lines are modified to have a delayed end to reionization. The HM12 model reionizes completely by $z = 6.7$, whilst the Late/Very Late models finish by $z = 6$, and the Delayed models finish at $z = 5.3$. These six histories allow us to bracket the range of possible reionization paths that are constrained by CMB and Lyα forest measurements. Right: the Thomson optical depth to electron scattering as a function of redshift for the different reionization histories. For comparison the Planck Collaboration (2018b) and Planck Collaboration (2016) CMB results are shown in green (with shading indicating the 1 σ uncertainty).

As in Chapter 3, we will test the three bracketing reionization histories first established in Choudhury et al. (2015), referred to as HM12, Late and Very Late. The evolution of the mass-averaged global neutral fraction in these models can be seen in the left panel of Figure 4.1 with the solid lines. We note that in the early HM12 model, reionization ends (i.e. when $\langle x_{\rm HII}(z_{\rm end})\rangle = 1$) at $z = 6.7$. The Late model is the same as the HM12 model but shifted in redshift so that reionization is completed by $z = 6$. Finally in the Very Late model reionization also ends at $z = 6$ but with a different redshift gradient $\mathrm{d}\langle x_{\rm HII}\rangle/\mathrm{d}z$, so that it finishes more abruptly. In this chapter we also consider modified versions of these three models in which the end of reionization is delayed to $z \sim 5.3$, but with the same evolution at higher redshifts. These will be referred to as the "Delayed" models (in comparison to the "Original" models). Physically these delayed models can result from an ionising emissivity evolution that peaks around $z \sim 7$ and then falls dramatically at lower redshifts, as suggested in Puchwein et al. (2019). We note that the Delayed Very Late reionization history has a similar neutral fraction evolution to the model of Kulkarni et al. (2019a), which was found to reproduce the opacity fluctuations in the Lyman-α forest (after reionization). The Delayed models are shown in Figure 4.1 with the dash-dotted lines. In the right panel of Figure 4.1 we compare the electron scattering optical depths predicted by these histories with the recent Planck CMB measurements. The low value of

$\tau = 0.054 \pm 0.007$ measured by Planck Collaboration (2018b) favours a later reionization, and we see that our Very Late models are consistent within 1σ of this recent measurement.

These reionization histories are then calibrated on the simulation so that the background photoionisation rate, $\Gamma_{\rm HI}$, (which is naturally coupled to the global average ionisation fraction) is self-consistent. The strength of the self-shielding and the equilibrium value of the neutral fraction within ionised regions is dependent on this UV background. Our calibrated simulation is also consistent with observed background photoionisation rates and ionising photon mean free paths (Faucher-Giguère et al., 2009; Calverley et al., 2011; Wyithe & Bolton, 2011) and the observed CMB optical depth Planck Collaboration (2018b). We note that our Very Late models are also consistent with current observational estimates from the SILVERRUSH survey on the neutral fraction evolution (see the additional material in Section 4.6.3).

4.2.2 Populating haloes with LBGs

We implement the Improved Conditional Luminosity Function (ICLF) method of Trenti et al. (2010) in order to populate the dark matter haloes in our simulations with LBG-type galaxies. This involves abundance matching the halo mass function to the observed UV luminosity function (as in, e.g. Lee et al., 2009; Bouwens et al., 2008). This assumes that each halo hosts one galaxy, and then equates the number of halos above a certain mass M_h with the number of LBGs above a certain luminosity $L_{\rm UV}$,

$$\epsilon_{\rm DC}(M_h, z) \int_{M_h}^{\infty} n(M, z) {\rm d}M = \int_{L_{\rm UV}}^{\infty} \phi(L, z) {\rm d}L, \tag{4.1}$$

where $n(M, z)$ is the halo mass function, $\phi(L, z)$ is the UV luminosity function, and $\epsilon_{\rm DC}(M_h, z) \leq 1$ is a mass and redshift dependent duty cycle which accounts for how likely it is that we will observe the galaxy hosted by a given halo at a particular time. Eq. (4.1) implicitly defines a mapping between halo mass and UV luminosity, $L_{\rm UV}(M_h)$. We use the same form of the duty cycle as Trenti et al. (2010),

$$\epsilon_{\rm DC}(M_h, z) = \frac{\int_{M_h}^{\infty} {\rm d}M \left[n(M, z) - n(M, z_{\Delta t}) \right]}{\int_{M_h}^{\infty} {\rm d}M \; n(M, z)}. \tag{4.2}$$

Note that,

$$\Delta t = t_H(z) - t_H(z_{\Delta t}), \tag{4.3}$$

where t_H is the local Hubble time (Trenti & Stiavelli, 2009). Δt is a free parameter in this duty cycle model; in this work we choose $\Delta t = 50$ Myr (originally a duty cycle with $\Delta t = 200$ Myr was employed by Trenti et al., 2010). The time interval between $z = 10$ and $z = 6$ (when reionization is underway) is less than 500 Myr, and so we choose a smaller Δt to reflect the bursty nature of star formation across this period. We note that other numerical work has found similar time scales for luminosity variation between 10–100 Myr, such as Rosdahl et al. (2018) who found large temporal variation of the fraction of escaping ionising photons, $f_{\rm esc}$, from galaxies during reionization. Our choice will be further

justified when we present the effect that varying Δt has on the clustering signal in Section 4.4.1. The $\Delta t = 50$ Myr duty cycle as a function of redshift and mass is shown in Figure 4.2. We finally note that the Δt parameter itself may in reality evolve with cosmic time, which would affect the evolution of the clustering signal of galaxies.

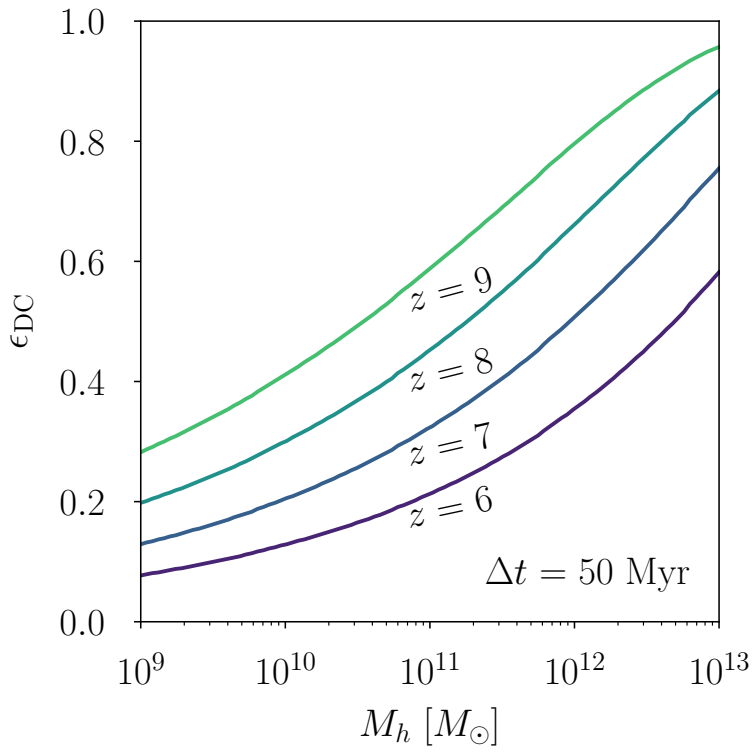


Fig. 4.2 Effective duty cycle defined by Trenti et al. (2010), as a function of halo mass and redshift, where Δt is a free parameter. This weights the abundance matching towards haloes that could have formed within Δt; for example shown here is our fiducial choice of $\Delta t = 50$ Myr.

A heuristic picture of this duty cycle is as follows: although we expect galaxies to exist in most dark matter haloes (above a minimum mass), it is not the case that we expect these galaxies to always be bright enough (in the UV) to be observed. The stochastic nature of star formation, and possibly geometric radiative transfer effects, might allow us to only observe a fraction of the underlying galaxy population (at a given time). In the above abundance matching procedure, enforcing a duty cycle will alter the mapping $L_{\rm UV}(M_h)$ in such a way as to shift the mapping to lower masses. This means lower mass haloes can be brighter (compared to the result if $\epsilon_{\rm DC} = 1$), which will further impact the clustering. Recently, Gurung-López et al. (2019) explored the role of radiative transfer effects within a given Lyα emitting galaxy, and found that carefully including these effects for a population of LAEs causes the LAEs to reside in less massive host haloes (compared to when such RT effects are neglected). Although our duty cycle is somewhat more agnostic to specific internal galaxy physics, it has the same effect. As in that work, we find that populating less massive haloes leads to better agreement with observables such as the clustering signal (Lee et al., 2009).

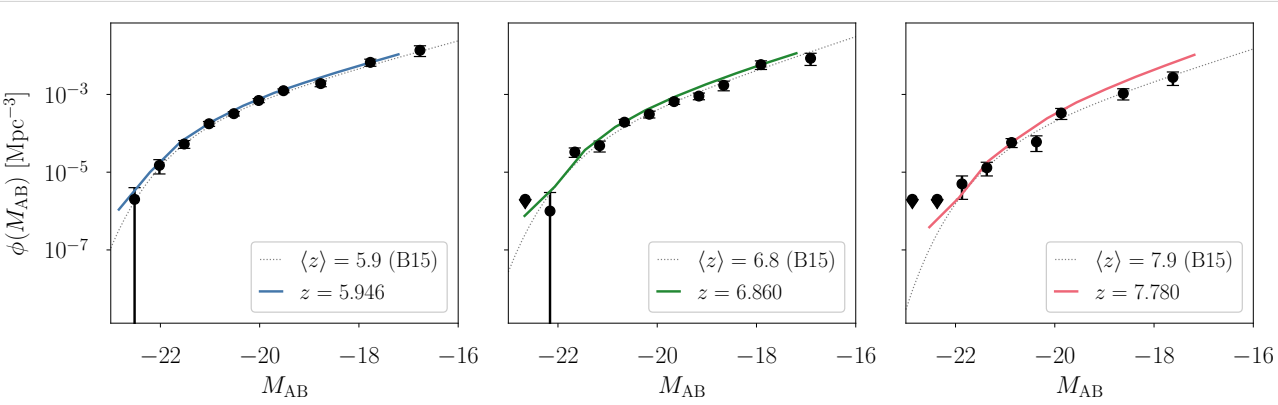


Fig. 4.3 The evolution of the LBG UV luminosity function. The thick solid lines show our model predictions, compared to the data-points from Bouwens et al. (2015) (and the dotted line best fit Schechter functions). The abundance matching is performed at $z = 5.9$, so the agreement in the left-most panel is by construction. At higher redshift the evolution is due to the evolution of the halo mass function, and we see good agreement between our model predictions and the observed luminosity functions.

We perform the abundance matching detailed in Eq. (4.1) using the Sheth-Mo-Tormen (SMT) halo mass function (Sheth et al., 2001), with HMFcalc (Murray et al., 2013). For the UV luminosity functions, we employ the best-fit Schechter functions from Bouwens et al. (2015). In particular we calibrate our luminosity mapping $L_{\rm UV}(M_h)$ at $\langle z\rangle = 5.9$, and can then apply this to the halo populations at each redshift of interest. Trenti et al. (2010) found that their mapping, calibrated at $z \sim 6$, was able to capture the evolution down to $z \sim 4$ of the observed UV luminosity function. The shape of $L_{\rm UV}(M_h)$ we find is similar to that found by Trenti et al. (2010), but shifted to lower masses due to the lower Δt that we employ in this work.

We can test that the evolution of the UV luminosity function is well matched by the underlying halo mass evolution by using the $L_{\rm UV}(M_h)$ mapping on the simulated halo populations. In Figure 4.3 we show this evolution starting with the calibrated redshift $z = 5.9$ on the left, and then higher redshifts on the right. Note the average redshift of the observational sample can be compared to the precise redshift of the simulation snapshot in the lower righthand subplot legends.

After this step in the framework, we now have an observationally calibrated mock sample of LBGs in our simulation. A subset of these will end up as a final LAE mock sample, after the selection detailed below.

4.2.3 LAE equivalent width distribution

We implement the rest-frame equivalent width (REW) distribution model proposed by Dijkstra & Wyithe (2012) in order to determine which of our LBG population will have observable Lyα emission. This is calibrated to empirical equivalent width distributions, therefore bypassing the need to model Lyα escape fractions and emission mechanisms. The model starts with a $M_{\rm UV}$-dependent distribution

of Lyα REWs,

$$P(\mathrm{REW}|M_{\mathrm{UV}}) = \mathcal{N} \exp\left(-\frac{\mathrm{REW}}{\mathrm{REW}_c(M_{\mathrm{UV}})}\right), \tag{4.4}$$

where $\mathrm{REW}_c(M_{\mathrm{UV}})$ is a characteristic REW for a given M_{UV}, given by (the best fit model of Dijkstra & Wyithe, 2012),

$$\mathrm{REW}_c(M_{\mathrm{UV}}) = 23 + 7(M_{\mathrm{UV}} + 21.9) + 6(z - 4). \tag{4.5}$$

The normalisation is defined such that the population has $\mathrm{REW}_{\mathrm{min}} \leq \mathrm{REW} \leq \mathrm{REW}_{\mathrm{max}}$, where we choose $\mathrm{REW}_{\mathrm{max}} = 300$ Å, and $\mathrm{REW}_{\mathrm{min}}$ is a function of M_{UV},

$$\mathrm{REW}_{\mathrm{min}} = \begin{cases} -20 \text{ Å} & M_{\mathrm{UV}} < -21.5; \\ 17.5 \text{ Å} & M_{\mathrm{UV}} > -19.0; \\ -20 + 6(M_{\mathrm{UV}} + 21.5)^2 \text{ Å} & \text{otherwise.} \end{cases} \tag{4.6}$$

Therefore the explicit form for the (M_{UV}-dependent) normalisation is,

$$\mathcal{N} = \frac{1}{\mathrm{REW}_c}\left(\exp\left(\frac{-\mathrm{REW}_{\mathrm{min}}}{\mathrm{REW}_c}\right) - \exp\left(\frac{-\mathrm{REW}_{\mathrm{max}}}{\mathrm{REW}_c}\right)\right)^{-1} \tag{4.7}$$

This choice of M_{UV} dependence reproduces the so-called Ando relation (Ando et al., 2007). We note Dijkstra & Wyithe (2012) compared this choice of distribution to the observed distributions at z = 3.1, 3.7 and 5.7 by Ouchi et al. (2008). They found good agreement at the lower redshifts, but that their model over-estimated the number of large REW systems at the higher redshifts. This was calculated in the absence of attenuation by the CGM/IGM, which is non-zero even after reionization has completed at $z \sim 6$. We find that the inclusion of an IGM transmission fraction < 1 compensates for the overprediction, and hence we do not modify the model to try to correct this. We further discuss this in Section 4.3.3.

For each of the haloes in our LBG sample, we draw a random REW from this conditional probability distribution, thereby assigning them a Lyα luminosity defined by,

$$L_{\mathrm{Ly}\alpha} = \frac{\nu_\alpha}{\lambda_\alpha}\left(\frac{\lambda_{\mathrm{UV}}}{\lambda_\alpha}\right)^{-\beta-2} \times \mathrm{REW} \times L_{\mathrm{UV},\nu}, \tag{4.8}$$

where $\nu_\alpha = 2.47 \times 10^{15}$ Hz is the Lyα transition frequency, $\lambda_\alpha = 1216$ Å the corresponding Lyα wavelength, $\lambda_{\mathrm{UV}} = 1600$ Å is the rest-frame wavelength at which the UV luminosity function was measured (Bouwens et al., 2015), $\beta = -1.7$ is the assumed UV spectral index, and the UV luminosity density, $L_{\mathrm{UV},\nu}$, is related to M_{UV} by (Ouchi et al., 2008),

$$M_{\mathrm{UV}} = -2.5 \log L_{\mathrm{UV},\nu} + 51.6. \tag{4.9}$$

Table 4.1 Observational selection thresholds[†] used in this work to generate mock observed samples.

Based on survey	z	$\mathrm{REW}_{\mathrm{min}}$ [Å]	$L^{\dagger\dagger}_{\mathrm{Ly}\alpha,\mathrm{min}}$ [erg/s]
Konno et al. (2018)	5.7	10	6.3×10^{42}
Konno et al. (2018)	6.6	14	7.9×10^{42}
Ota et al. (2017) Itoh et al. (2018)	7.0	10	2×10^{42}
Konno et al. (2014)	7.3	0	2.4×10^{42}

Note:
[†]The observational surveys we derived these limits from measured different fields on the sky, across which different selection thresholds were sometimes applied. The values we have chosen are representative despite this variation.
[††]As in Dijkstra & Wyithe (2012), when not quoted in the original survey, we estimate L_{min} as the lower bin edge of the lowest luminosity bin in the presented luminosity function.

Having generated $L_{\mathrm{Ly}\alpha}$ for the LBG mock sample, we then apply selections based on luminosity and equivalent width limits to match a given observational study. Table (4.1) shows some of the observational thresholds used in the SILVERRUSH survey (Ouchi et al., 2018; Konno et al., 2018; Shibuya et al., 2018), which we adopt here.

We note that Dijkstra & Wyithe (2012) concluded that equivalent width and luminosity cuts are only an approximation to the real selection thresholds used in observational LAE studies. They found that in order to match both the observed equivalent width distribution and the luminosity function, they had to scale the number density by ~ 0.5. With this scaling included, their modelling then matched the redshift evolution very well.

When calculating the IGM transmission (as described in detail in section 4.2.4) in our simulations we find that even at $z = 5.7$ the transmission redwards of Lyα is not 100%. As was suggested in Sadoun et al. (2017) and further explored in Chapter 3, this is due mostly to the neutral gas in the outer part of the host halo (close to the virial radius), controlled by the background photoionisation rate (see also Section 4.3.2). In our delayed models there is also an increased fraction of neutral gas in the rest of the IGM, but we find the attenuation at these redshifts is dominated by CGM and surrounding halo gas[1]. This means that when we include the transmission we find the number density of 'observed' LAEs drops, in good agreement with the real observed luminosity functions. We therefore choose to use the REW distribution from Dijkstra & Wyithe (2012) to model the 'intrinsic' distribution of our population, which is then attenuated by the IGM to give the 'observed' REWs and Lyα luminosities. As such we do not apply the ~ 0.5 scaling to the number density in our model predictions which was needed in Dijkstra & Wyithe (2012); we find agreement with the observational results without this.

[1]In this thesis we will refer to the infalling gas surrounding the halo as the CGM gas. We note however that we include gas out to larger scales ($r \lesssim 10R_{\mathrm{vir}}$) than the more common observational definition of the CGM gas (at scales closer to R_{vir}, e.g. Steidel et al., 2010)

This means we now have a mock LAE sample, in which each object has a value of M_h, $M_{\rm UV}$, REW, and $L_{{\rm Ly}\alpha}$ that conforms to the chosen observational selection window. The final step in the framework, before we can compare with observations, is to calculate the IGM transmission fraction for a given halo in the sample, and reduce the luminosity accordingly.

4.2.4 Lyα transmission

We calculate the CGM/IGM transmission as in Chapter 3, extracting sightlines[2] through the halo sample to find the optical depth to Lyα, $\tau_{{\rm Ly}\alpha}(v)$, as a function of velocity offset from the emitter (Bolton & Haehnelt, 2007). This can be used to calculate a transmission fraction, assuming an intrinsic emission profile $J(v)$ for the galaxy, which itself accounts for the radiative transfer within the halo as photons escape the galaxy's ISM. The emission profiles of LAEs seen in both observations (e.g. Hashimoto et al., 2015) and radiative transfer simulations (e.g. Zheng & Wallace, 2014) are non-trivial to model, with complicated dependences on the local gas dynamics.

As in Chapter 3 we make the simplifying assumption that the emission profile is a single-peaked Gaussian profile with width σ_v, offset redwards from the systemic by Δv due to resonant scattering within the halo. We choose $\sigma_v = 88$ km/s as in Choudhury et al. (2015). Our fiducial choice for the velocity offset is to set it proportional to the virial circular velocity of a given LAE's host halo, $\Delta v \propto v_{\rm circ}$. This is motivated by observational and theoretical work such as Verhamme et al. (2018); Smith et al. (2018); Dijkstra et al. (2006); Neufeld (1990), which has shown that the radiative transfer (as a strong function of HI opacity) in a galaxy's ISM/CGM leads to a coupling between the dispersion of the Lyα line and its velocity offset. We might expect the dispersion to be proportional to a halo's circular velocity, which in turn gives the same proportionality for the offset. Previously in Chapter 3 we used a fixed value of $\Delta v = 100$ km/s, neglecting any dependence on the emitter properties and assuming no variation across the population. Although this choice was simplistic, we found that the relative transmission fraction, $T^{\rm IGM}_{{\rm Ly}\alpha}(z)/T^{\rm IGM}_{{\rm Ly}\alpha}(z = 5.7)$, was largely insensitive to this choice. In the present chapter we now need to consider the absolute transmission fraction at a given redshift and so we have updated our model for the intrinsic emission.

Other theoretical works have treated the intrinsic emission profile differently, for example Inoue et al. (2018) used the single peaked profiles that result from full radiative transfer calculations applied to an outflowing spherical volume of gas. This is strongly dependent on the chosen HI column density and outflow velocity; our modelling assumes an explicit halo mass dependence for the profile rather than fixing these quantities across the population.

We finally note that this is a poorly constrained quantity observationally, and so we later test the effect of varying our Δv assumption. In reality the intrinsic emission profile of an LAE will evolve and vary across its lifetime, a feature we do not attempt to model in this work.

[2] This includes a 20 cMpc/h region around the halo with a higher resolution of 9.8 ckpc/h.

The transmission fraction for this profile can then be calculated as (Mesinger et al., 2015),

$$T_{\mathrm{Ly}\alpha}^{\mathrm{IGM}} = \frac{\int \mathrm{d}v\, J(v)\, e^{-\tau(v)}}{\int \mathrm{d}v\, J(v)}. \tag{4.10}$$

Note that specifically this is the transmission fraction redwards of systemic; in choosing a single-peaked profile we have accepted that the IGM is sufficiently optically thick at the redshifts we consider such that even if a blue peak emerges after radiative transfer in the ISM, the HI damping wing of the IGM will not transmit bluewards of systemic. This method of separating the radiative transfer into a galaxy stage (which we simply model with our velocity offset Δv) and an IGM stage (which we model using the $e^{-\tau}$ approximation) has been employed successfully before, for example in Laursen et al. (2019, 2011). As we account for the galactic radiative transfer already, we exclude the host halo gas in our calculation of $\tau_{\mathrm{Ly}\alpha}$. Our fiducial choice is to exclude halo gas within 1 R_{vir} of the halo centre, the effect of which was tested in Chapter 3. Similarly Laursen et al. (2019) chose to transition between the two calculation regimes at 1.5 R_{vir}. We reiterate that the choice of intrinsic emission profile has a strong effect on the transmission fraction, and therefore also on the resulting population statistics for our LAE mock catalogues.

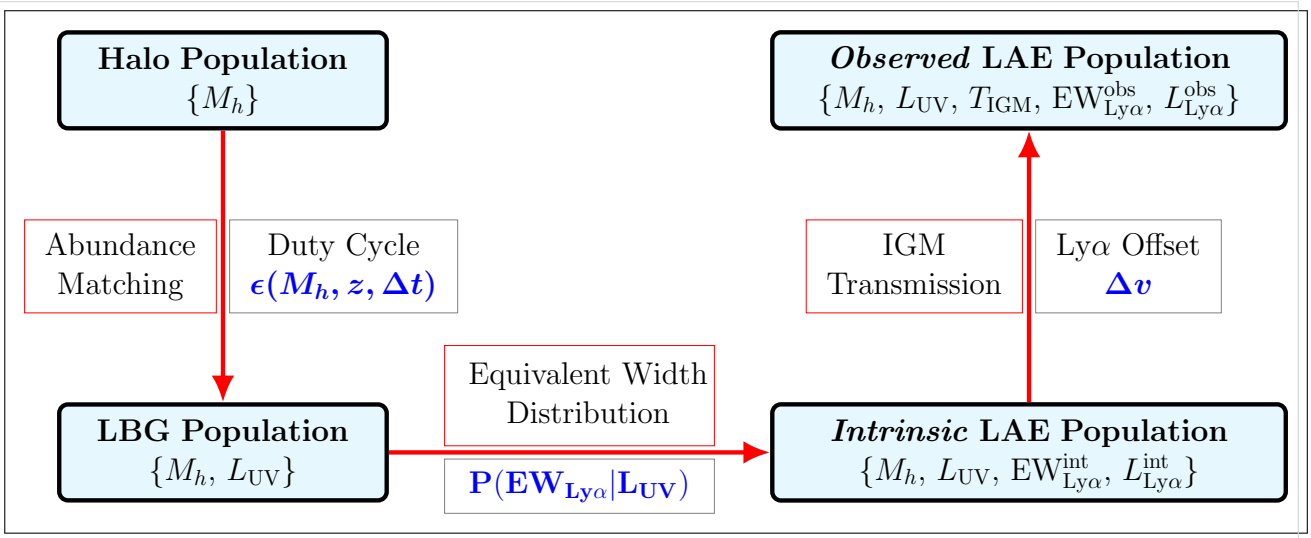


Fig. 4.4 Summary schematic representing the different parts of the LAE modelling: first the halo population of a given simulation snapshot is abundance matched to the observed UV population with an assumed duty cycle, to create a mock LBG population. Then we draw from an empirical 'intrinsic' REW distribution to create a mock 'intrinsic' LAE population. Finally the effects of IGM transmission are taken into account to create a mock 'observed' LAE population.

Having performed the transmission fraction calculation for each halo, we can update the derived Lyα luminosity,

$$L_{\mathrm{Ly}\alpha}^{\mathrm{obs}} = T_{\mathrm{Ly}\alpha}^{\mathrm{IGM}} L_{\mathrm{Ly}\alpha}, \tag{4.11}$$

and similarly for the equivalent widths. This completes our generation of a mock LAE sample from our simulated halo population. This framework allows us to generate samples at any desired redshift to compare with observations, and make predictions at higher redshifts for the evolution of the luminosity function and clustering signal. Note that when the transmission fraction falls below unity, some of the LAEs will drop below the flux-limit, hence we reapply the selection after accounting for transmission. In Figure 4.4 we show a summary schematic for the different stages of the LAE modelling.

4.2.5 Modelling caveats

Although this model framework is successful at matching the evolution of observed luminosity functions and clustering studies (as we will present in Section 4.3), there are a number of important caveats to consider:

- We do not use sub-find catalogs with satellite haloes. As we are only using central haloes we will naturally underestimate the 1-halo term of the clustering signal. However at the redshifts of interest $z \gtrsim 6$ the halo occupation distribution (HOD) populates only very large mass haloes with satellites. For example Bhowmick et al. (2018) found that the satellite fraction of haloes is around ~ 10% at $z = 7.5$, and that the mean number of satellites is less than unity for halo masses below $M_h \sim 10^{11}\, M_\odot/h$. Similarly Gurung-López et al. (2019) found that satellite galaxies only start to dominate the abundance of haloes with mass $M_h \gtrsim 10^{12}\, M_\odot/h$.

- There are a number of tunable parameters in this model. The primary free parameters are Δt which controls the duty cycle, and Δv which controls the 'intrinsic' emission profile and hence has a strong effect on the calculated transmission fraction. We take the best-fit REW distribution from Dijkstra & Wyithe (2012) and so do not leave any of those model parameters free, but use their empirically constrained values.

- Although we predict the transmission, and leave the duty cycle as a free parameter, these variables have a degenerate effect on the clustering. Increasing (decreasing) the duty cycle (the transmission) can lead to an increase in the measured clustering signal. We have used physically motivated values for the free parameters Δt and Δv.

4.3 Results

We now discuss the results of applying our LAE framework to the halo population in the Sherwood simulations. In particular we consider the IGM transmission across this population, and how this affects the luminosity function evolution. We also confirm that our model matches the observed equivalent width distribution and luminosity function at $z = 5.7$. Taking this redshift as an anchor we then also create mock survey slices from which we calculate the angular correlation function, a

2-point measure of the clustering signal. Finally we extend our models to higher redshifts to make predictions for future surveys.

Note for all plots where we compare our model predictions to the observational data, we have taken slices of the simulation volume and calculated the relevant statistical quantity for the sample of LAEs in each slice. The shading shown for the model predictions corresponds to 68% scatter across the slices. Specifically we divide the box up in configuration-space into 10 slices perpendicular to the direction along which we calculated the transmission, giving a comoving thickness of 32 cMpc/h. This is not exactly equivalent to the narrowband selection, which is instead a slice in redshift (velocity) space; however given the width of the narrowband slice and the comparatively small amplitudes of peculiar motions of the LAEs in our simulation volume, we find that the results using configuration-space slicing are indistinguishable from the velocity-space slicing. In all figures we show the intrinsic (unattenuated) quantities using a dashed grey line, whilst the different model (attenuated) quantities are shown using coloured solid lines (original reionization histories) and dash-dotted lines (delayed-end models).

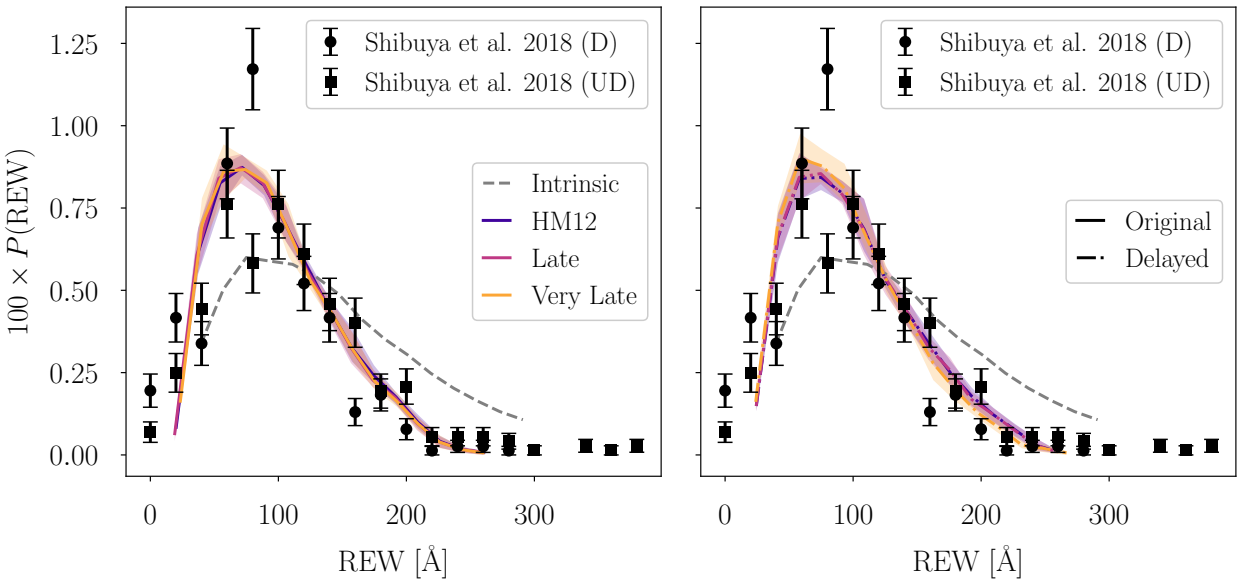


Fig. 4.5 Comparison of our model mock REW distribution at $z = 5.756$ with the observed distribution from the SILVERRUSH survey at $z = 5.7$. The different reionization histories are shown by the different colours, whilst the intrinsic un-attenuated model is shown with the grey dashed line. On the left hand panel in the solid lines we show the original reionization histories; on the right hand panel we show the Delayed models with dash-dotted lines. For the chosen emission offset parameter, $\Delta v \propto v_{\rm circ}$, we find that the IGM transmission corrects the over-abundance of high REW systems predicted by the intrinsic model.

4.3.1 IGM attenuation of the $z = 5.7$ equivalent width distribution

In Figure 4.5 we compare the observed REW distribution from the SILVERRUSH survey (Shibuya et al., 2018) at $z = 5.7$ with that predicted from our mock LAE population. We show the un-attenuated 'intrinsic' distribution with the dashed grey line, whilst the IGM attenuated distributions are shown for the three different reionization histories in blue (HM12), purple (Late) and orange (Very Late).

We see at this redshift that although our original reionization histories give a global average ionised fraction of unity, the IGM transmission fraction is sufficiently below unity that it has a significant effect on the observed REW distribution. This attenuation results from (photoionisation) equilibrium and self-shielded neutral gas around haloes which contributes little to the average ionised fraction but has a strong effect on the Lyα attenuation (see Section 4.3.2). In particular the transmission fraction distribution is such that it attenuates the high REW objects, thereby reducing the over-abundance of such objects which is predicted by the intrinsic distribution. We have chosen to use a velocity offset of $\Delta v = a\, v_{\rm circ}$ where,

$$a = \begin{cases} 1.5 & \text{Original,} \\ 1.8 & \text{Delayed,} \end{cases} \tag{4.12}$$

which gives a distribution consistent with the observed data. We also tested proportionality constants $a = 1$ and $a = 2$ and found that these resulted in either too much attenuation or too little, respectively. The larger value of a is required in the Delayed models, where reionization has not yet ended by $z = 5.7$, so there is considerably more attenuation by residual neutral gas. In particular we find that these models require a lower background photoionisation rate, therefore increasing the amount of neutral gas present in the outer parts of the LAE host haloes (in the CGM). The presence of this gas is sufficient to reduce the transmission redwards of Lyα.

The resonant scattering of Lyα radiation by neutral hydrogen within the galaxy tends to diffuse the emission profile away from line-centre, with radiation escaping in blue or red peaks where the scattering cross-section is smaller. If the galaxy has outflows, the red peak can be enhanced such that the dominant emission comes at redder velocities. This is seen in shell models with an expanding H I outflow (e.g. Dijkstra et al., 2011; Verhamme et al., 2006). We might therefore expect that the velocity offset Δv should be coupled to the galaxy wind velocity which, to avoid stalling, must be of order the escape velocity of the halo, i.e. $\sim \sqrt{2} v_{\rm circ}$. Hence the values of a in Eq. (4.12) are reasonable.

4.3.2 Attenuation from the partially neutral CGM

As seen in Section 4.3.1, even at $z = 5.756$ (where our original reionization histories are fully reionized, i.e. $\langle x_{\rm HII} \rangle_{\rm M} = 1$) there is still an attenuation of radiation redwards of Lyα. This is due to infalling neutral gas around the halo (including the CGM around the LAEs), which is not fully ionised by either the LAE itself or the ionising UV background. Sadoun et al. (2017) found that the drop in observed numbers of LAEs doesn't necessarily imply a largely neutral IGM, since this infalling CGM

gas can also bring the Lyα transmission below 100%. We note again that our usage of the term CGM refers to a larger volume of the infalling gas that surrounds the host halo than the more common observational definition.

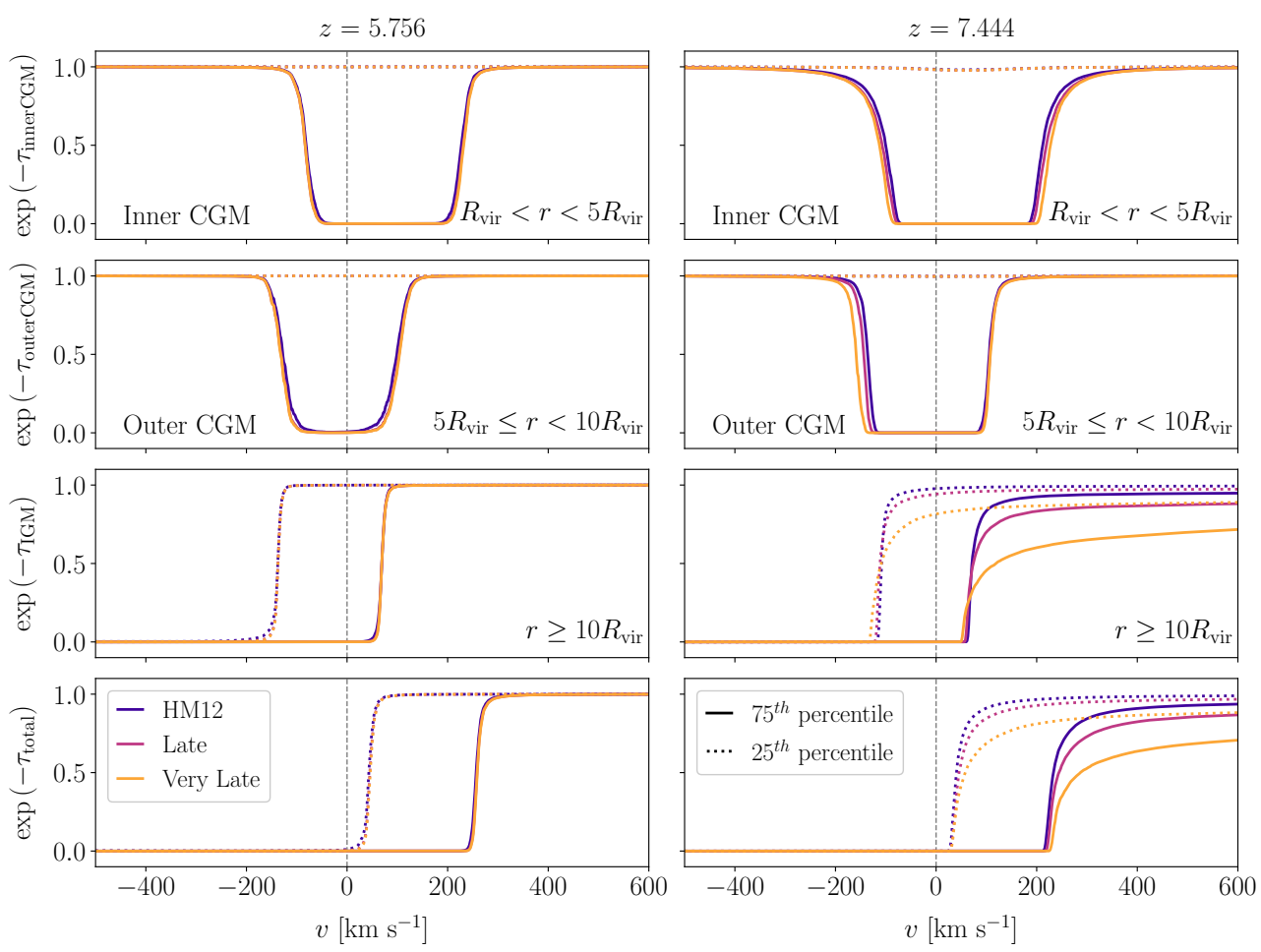


Fig. 4.6 Separating the contributions to the Lyα transmission: (i) in the top panels we show the transmission due to the CGM and surrounding gas (split into an "inner" part between $R_{\rm vir} < r < 5R_{\rm vir}$ and an "outer" part between $5R_{\rm vir} \leq r < 10R_{\rm vir}$) (ii) in the middle panels the contribution from the IGM ($r \geq 10R_{\rm vir}$); (iii) in the bottom panels the transmission from all the gas outside the halo ($r > R_{\rm vir}$). In all panels the solid (dotted) curves correspond to the 75^{th} (25^{th}) percentile as measured across our mock observed sample of LAEs (i.e. after $L_{\rm Ly\alpha}$ and REW selection), which spans an order of magnitude in luminosity. The transmission is shown as a function of velocity offset from line-centre, at redshift $z = 5.756$ on the left and $z = 7.444$ on the right. The three original reionization histories are shown using the coloured solid lines. For clarity we do not plot the delayed-end models, however their corresponding transmission curves are very similar.

To quantify this further, in Figure 4.6 we plot the 75^{th} and 25^{th} percentiles of the transmission along sightlines to our observed samples of mock LAEs. We show the transmission as a function of velocity offset from line-centre due to: (i) the infalling CGM and surrounding gas (split into an "inner"

part between $R_{\rm vir} < r < 5R_{\rm vir}$ and an "outer" part between $5R_{\rm vir} \leq r < 10R_{\rm vir}$ around the halo centre); (ii) the exterior IGM gas ($r \geq 10R_{\rm vir}$); and (iii) the total gas around the LAE. We show this for our three original reionization histories at the two bracketing redshifts[3]. We see that there is significant halo-to-halo variation in the CGM component, where the transmission can vary from 100% at all velocities (in the 25^{th} percentile) to 0% around $v = 0$ km/s (in the 75^{th} percentile). In the left panel (at $z = 5.756$) we see that the IGM is transmitting at ~ 100% redwards of Lyα (i.e. $v > 0$ km/s) for the 25^{th} percentile, whereas the 75^{th} percentile starts to transmit at ~ 100% for redder velocities $v \gtrsim 100$ km/s. Furthermore we see the drop in transmission due to the CGM gas extends redwards of line-centre, because the gas is infalling onto the halo. In particular the "inner" part of the CGM attenuates redwards of the "outer" part because the amplitude of the infalling gas velocity peaks in that region. This means that radiation redwards of line-centre can be blue-shifted in the frame of the gas towards line-centre, and hence resonantly scattered out of the line of sight. The CGM transmission evolves across the redshifts as a function of the photoionisation rate, which controls how neutral the gas is.

In comparison the IGM transmission gradually decreases with increasing redshift as the average neutral fraction increases. Considering the shape of the attenuation imprinted by these different components, we note that the CGM evolution (dependent on the photoionisation rate) causes a velocity shift in the transmission curve along the horizontal axis, whilst the IGM evolution (a function of the average neutral fraction) causes less transmission, i.e. a shift along the vertical axis. In particular we note that near the end of reionization the horizontal shift caused by the CGM is the dominant component of the attenuation. At higher redshifts we can distinguish the different reionization histories because their average neutral fractions diverge significantly, causing varying amounts of vertical shift in the transmission curve. We find there is a luminosity (or mass) dependence in the evolution of these two components, explored further in the additional material in Section 4.6.2.

4.3.3 Lyα luminosity function evolution

Given the mock LAE sample for each redshift which includes the IGM transmission fraction and the intrinsic luminosity, we can construct the observed sample using Eq. (4.11). This allows us to construct the (spatial) luminosity function and compare to observed results. In Figure 4.7 we show the evolution of the luminosity function for our six reionization histories, using our fiducial $\Delta v \propto v_{\rm circ}$ model. From left to right the redshift of our mock population increases from $z = 5.756$ to $z = 7.444$. As described above, the luminosity function is calculated by slicing the simulation volume and taking the mean across the slices for each luminosity bin. The shading represents the 68% scatter around this slice mean.

We see that the reionization history which qualitatively fits the observed data across the available narrowband redshifts is the Delayed Very Late model, suggesting that a later reionization is most consistent (as found in Kulkarni et al., 2019a; Planck Collaboration, 2018b). We note that the most

[3]The transmission curves for the delayed-end models are similar, but are not shown to aid the clarity of Figure 4.6.

difficult constraint to match is the strong attenuation seen in the $z = 7.3$ data. The earlier reionization histories under-predict the IGM attenuation required to match the data at $z = 7.3$. Since these datasets require the deepest observations in order to find the very rare LAEs visible at such high redshifts, it is also possible that some of the $z = 7.3$ bins are not fully complete and may move up in the future SILVERRUSH data release. Attenuation at $z = 7.3$ may be evidence for an even later start to reionization, even later perhaps than our Very Late model. We also note that there is some inconsistency between the different observed datasets at $z = 7.0$, which has been discussed in the literature.

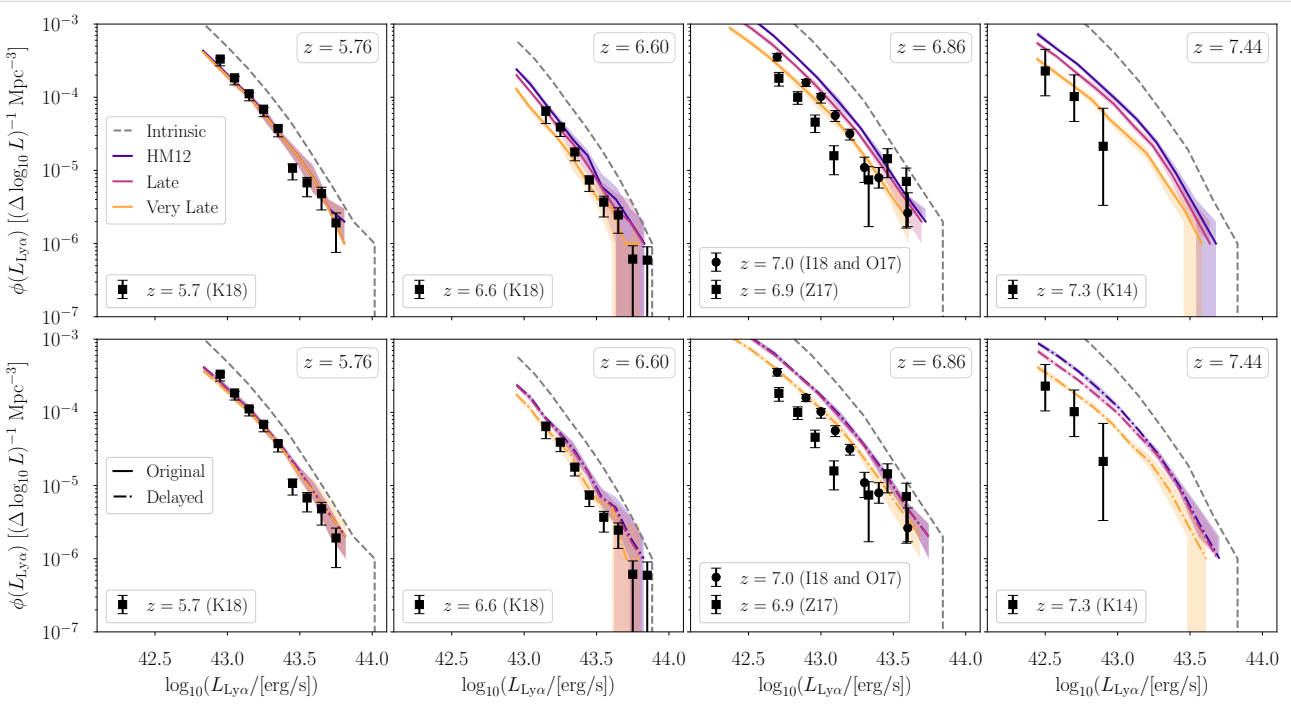


Fig. 4.7 From left to right: The luminosity function evolution, at $z = 5.756$, 6.604, 6.860 and 7.444, for each of the reionization histories shown as coloured lines. The top panel with solid lines shows the original three reionization histories, whilst the bottom panel with dash-dotted lines shows the delayed-end histories. Some observed data is overplotted for comparison with black markers, from Konno et al. (2018) (K18, $z = 5.7$ & 6.6), Itoh et al. (2018) (I18, $z = 7.0$), Ota et al. (2017) (O17, $z = 7.0$), Zheng et al. (2017) (Z17, $z = 6.9$), Konno et al. (2014) (K14, $z = 7.3$).

4.3.4 Clustering evolution

As with the luminosity function calculation, we divide the simulation volume into slices of approximately the same depth as the narrowband surveys (10 slices of depth 32 cMpc/h), and assume the same luminosity cuts. We then use the Landy & Szalay (1993) estimator,

$$w(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)}, \tag{4.13}$$

to calculate the angular correlation function $w(\theta)$, where $DD(\theta)$ is the number of galaxy-galaxy pairs at angular separation θ, $RD(\theta)$ the number of random-galaxy pairs, and $RR(\theta)$ the random-random pairs, all of which are normalised appropriately. We employ the swot code (Coupon et al., 2012) to perform the calculation efficiently. Our random field is generated by drawing from a uniform distribution, with similar number density to that of Ouchi et al. (2018) (see Coupon et al., 2018). In Figure 4.8 we plot the angular correlation function for each test redshift, showing the different reionization histories as in Figure 4.7. The scatter across the slices is shown by the shading, whilst the lines are the mean value of $w(\theta)$. We note that at both $z = 5.756$ (left) and $z = 6.604$ (middle-left) our predictions are within the scatter of the observational results from Ouchi et al. (2018). In the $z = 6.604$ panel, we already start to see the effect of the different transmission fractions predicted by the reionization history models. The Very Late model, with the lowest average transmission, gives the highest clustering signal.

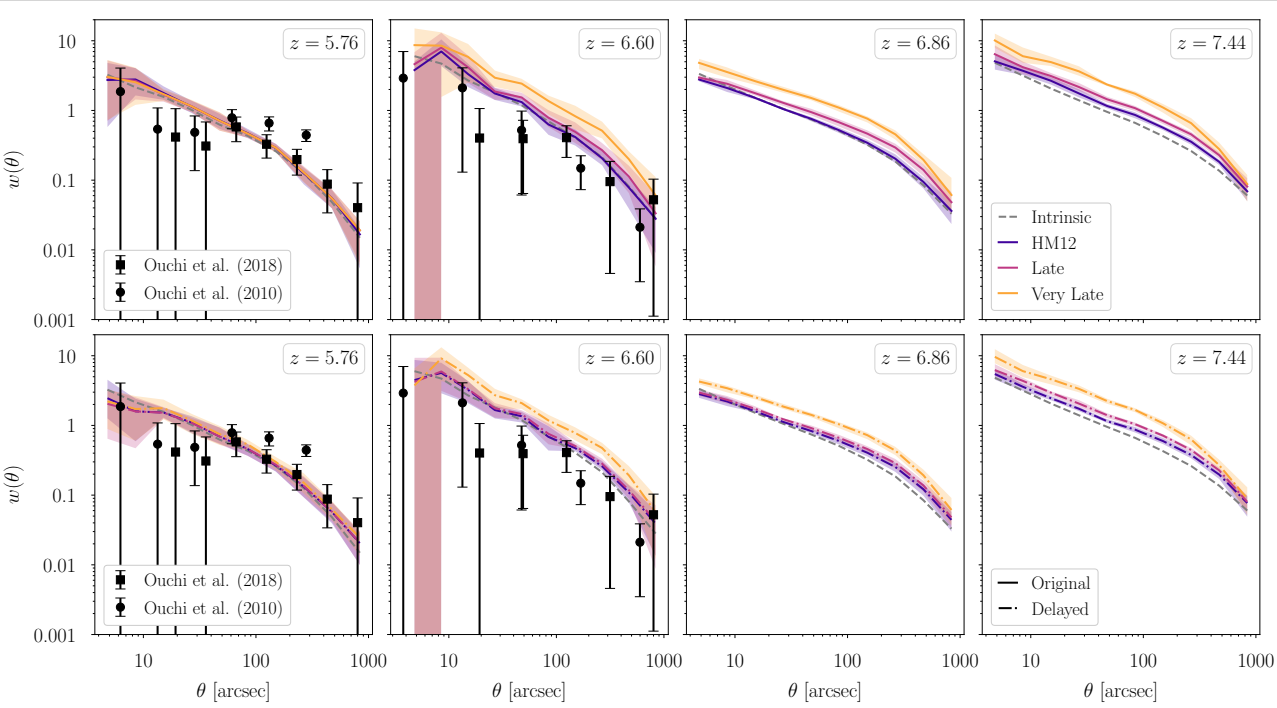


Fig. 4.8 From left to right: The angular correlation function, measuring the clustering signal at $z = 5.756$, 6.604, 6.860 and 7.444, for our fiducial model with $\Delta t = 50$ Myr. As in Figure 6, the top panel shows the original reionization histories in solid lines, whilst the bottom panel shows the delayed-end versions in dash-dotted lines. At $z = 5.756$ the reionization histories have converged (reionization has ended) and so the models are equivalent. At higher redshifts we start to see some divergence of the amount of clustering predicted for each reionization history, with the Very Late model predicting the most clustering power. For each redshift these mock surveys correspond to 10 slices of our simulation volume (each with area 320×320 (cMpc/h)^2), therefore simulating a total survey area of 2.2 Gpc^2. We also refer the reader to Table (4.1) for the mock selection limits used at each redshift.

Considering the SILVERRUSH clustering data (Ouchi et al., 2018, 2010) alone, there is very little evolution in the angular correlation function. Since we expect the clustering to be increased due both to the higher bias at higher redshifts as well as due to the ionised bubble structure of the IGM, this lack of evolution between z = 5.7 and 6.6 is puzzling, perhaps suggesting that samples are not yet large enough for an accurate determination of the clustering of this higher redshift. From the modelling perspective, it would be possible to reduce the predicted clustering signal at z = 6.6 further by using a shorter duty cycle, i.e. a lower value of Δt, however this will also affect the quality of the agreement at z = 5.7.

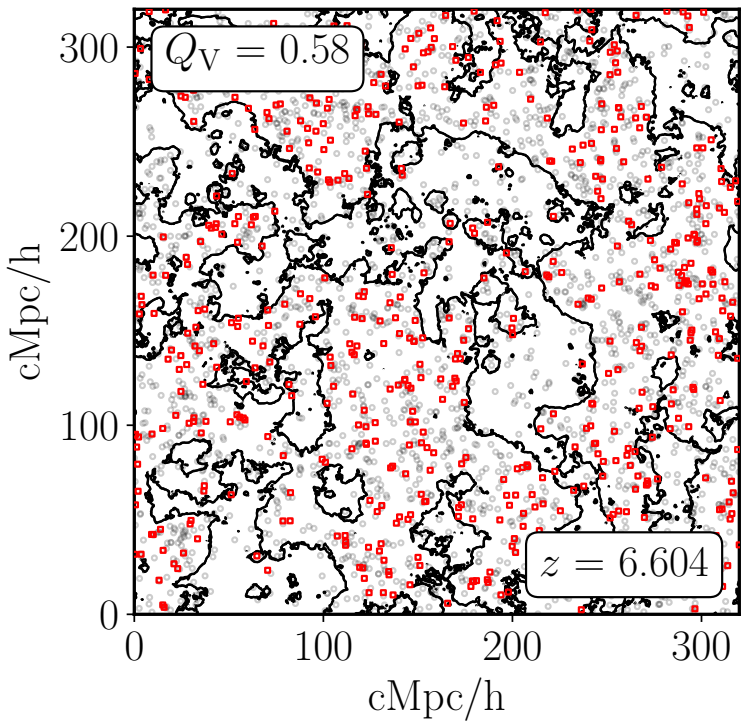


Fig. 4.9 A mock map of a sample of LAEs within a 32 cMpc/h slice of the simulation volume, showing the whole population in grey empty circles and those above the observational thresholds in red empty squares. A black contour at a value of half the maximum projected neutral fraction is also plotted, segregating the regions within this projected slice which are largely neutral from those which are ionised. It is visible by eye that the majority of the observed mock LAEs (in red) lie within the ionised regions.

4.3.5 Clustering predictions for $z \geq 7.0$

In the right-hand two panels of Figure 4.8 we make predictions for the clustering signal at redshifts z = 6.860 and = 7.444, which are close to the narrowband filters NB973 and NB101. We see a similar pattern as was observed for the left-hand panels: the clustering signal increases for all models, and in particular the HM12 history (in blue) exhibits the least clustering whilst the Very Late model (in orange) exhibits the most. On intermediary scales these models are non-overlapping at the 68%

scatter level. In particular the 2-point correlation function is a function of the reionization history, suggesting that measurements at these redshifts could be strongly constraining.

Apart from the angular correlation function, the clustering signal might also be useful for understanding ionised bubble structure deep into reionization (e.g. Kakiichi et al., 2016). As an example of how this might be possible, we plot a projected mock LAE slice in Figure 4.9 indicating the intrinsic population in grey and the observed population in red, at $z = 6.604$ for the Very Late reionization history. We also plot a contour partitioning the map based on the projected ionisation fraction, to indicate where the LAEs reside with respect to the ionised bubbles. We note that the observed LAEs seem to lie within ionised bubbles, whereas the unobserved objects are in neutral regions. This configuration could therefore allow us to constrain bubble sizes, perhaps in concert with proposed 21 cm methods (Giri et al., 2018a). We leave the construction of such methods to future work.

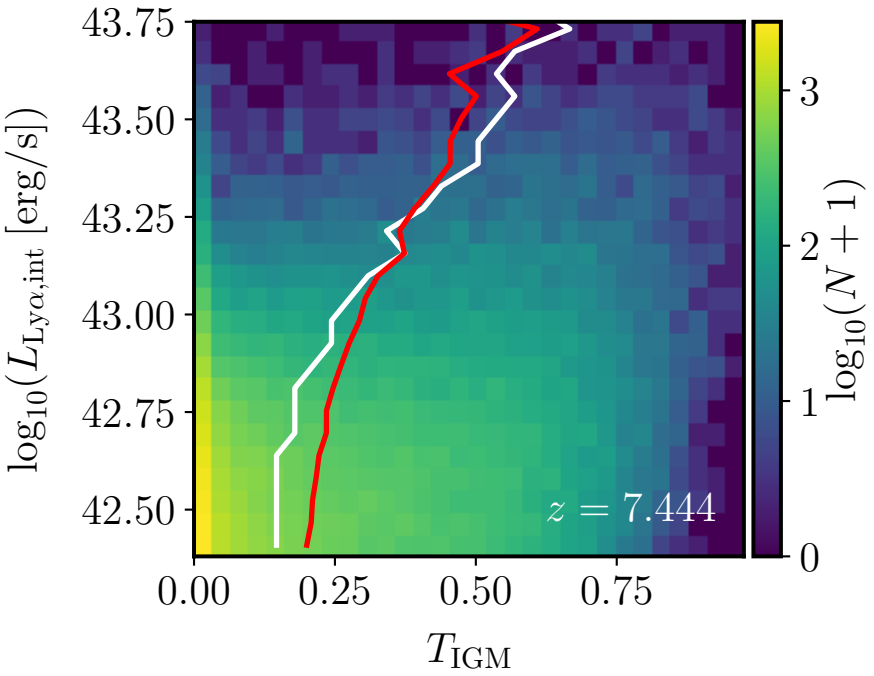


Fig. 4.10 Histogram for the transmission fraction distribution and the 'intrinsic' Lyα luminosity for our mock LAE population at $z = 7.444$. The number of objects in each bin is indicated by the colourmap, whilst the red (white) line indicates the mean (median) transmission fraction for a given luminosity bin. We see that the brighter LAEs are preferentially more visible than the fainter LAEs.

4.3.6 Differential evolution of the bright and faint end of the luminosity function

In Figure 4.10 we plot a 2D histogram for the mock LAE population at $z = 7.444$, binning by transmission fraction and 'intrinsic' Lyα luminosity. We see that for our chosen transmission model

the brighter LAEs are preferentially more visible than the fainter objects. In particular the mean and median transmission fractions are shown with the red and white lines, and we see that both curve towards higher transmission fractions as the luminosity increases.

Considering the results of Chapter 3, there are two components of this differential visibility to understand. Firstly the absolute transmission fraction at a given redshift is strongly dependent on the emission profile; for our models it is therefore dependent on the choice of Δv. Our choice in this work assumes $\Delta v \propto v_{\rm circ}$ which means that $\Delta v \propto M_{\rm vir}^{\frac{1}{3}}$. This partly explains the behaviour seen in Figure 4.10; other theoretical work such as Mason et al. (2018b) assumed $\Delta v \propto M_{\rm vir}$ and found a similar boost in the transmission of bright LAEs. However in Chapter 3 it was found that there can also be a differential visibility for the relative transmission fraction (i.e. the transmission fraction at a given redshift relative to another redshift). The differential evolution of this relative transmission was partly caused by the presence of brighter LAEs within larger ionised regions, as well as the different host environments of the haloes that the LAEs resided in. Importantly it was found that this differential evolution of the relative transmission was insensitive to the choice of Δv.

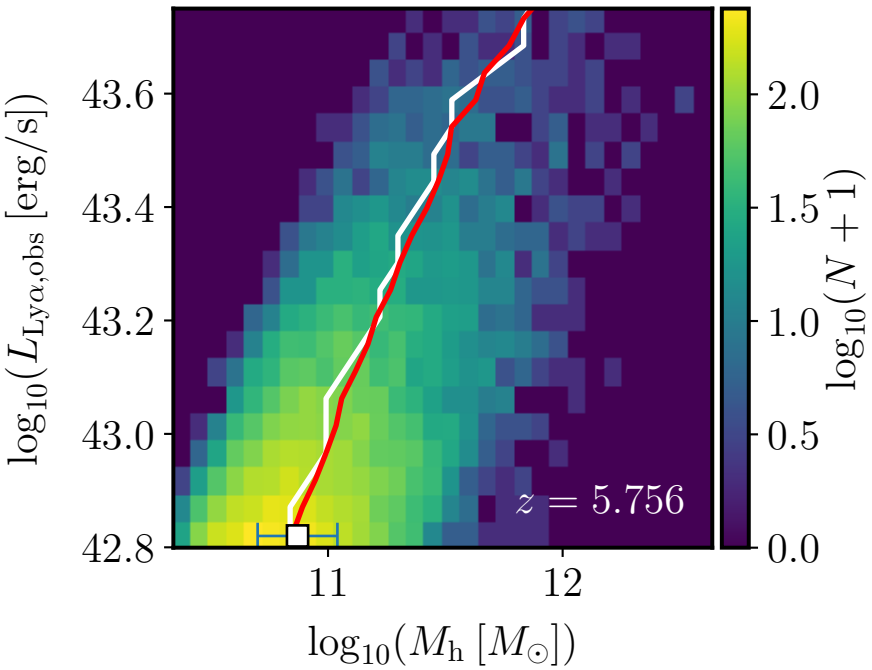


Fig. 4.11 Histogram for the mass distribution and the 'observed' Lyα luminosity for our mock LAE population at z = 5.756. The number of objects in each bin is indicated by the colourmap, whilst the red (white) line indicates the mean (median) host halo mass for a given luminosity bin. The white datapoint shows the median luminosity and effective halo mass for the NB816 sample of Khostovan et al. (2019).

4.3.7 Typical LAE masses

In Figure 4.11 we plot the distribution of host halo masses as a function of 'observed' Lyα luminosity for our mock population at $z = 5.756$. Overplotted on our model histogram we show the observed NB816 data from Khostovan et al. (2019) with a white marker. As in Figure 4.10 the mean and median of the distributions are shown in red and white lines. We see the strong correlation between halo mass and luminosity, expected from Eq (4.1), which prevails even after IGM attenuation. The mean host halo mass of our $z = 5.756$ mock sample is $10^{11.0} M_\odot$, whilst the minimum is $10^{10.3} M_\odot$ and the maximum is $10^{12.6} M_\odot$.

4.4 Discussion

Our model, detailed above, reproduces the evolution of the LAE luminosity function and angular correlation function reasonably well. The main free parameters were chosen in our fiducial model as $\Delta v = 1.5 v_{\rm circ}$ and $\Delta t = 50$ Myr. We now discuss further the motivation for these choices, and the effect of varying these parameters on the observables.

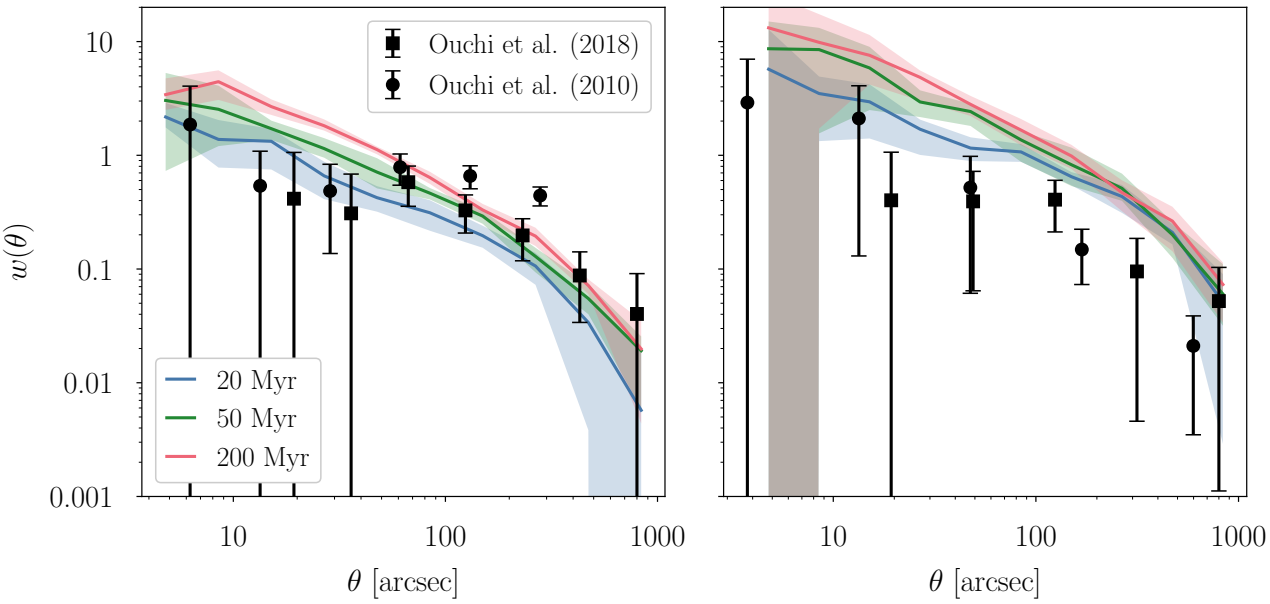


Fig. 4.12 Angular correlation functions at $z = 5.756$ (left) and 6.604 (right), comparing different choices of the Δt duty cycle parameter: a smaller value of 20 Myr in blue, our fiducial value of 50 Myr in green, and the fiducial value of Trenti et al. (2010) in red. All lines are calculated for the same reionization history.

4.4.1 The effect of varying Δt on the clustering

In Figure 4.12 we show the the angular correlation function at $z = 5.756$ and 6.604 for three different values of Δt: 20, 50 and 200 Myr. We see that increasing this parameter causes an increase in the

clustering power, especially at smaller scales. In particular comparing to observations from Ouchi et al. (2018, 2010), the 200 Myr duty cycle causes too much small scale correlation to be consistent with the observed correlation at scales around $10 \lesssim \theta \lesssim 60$ arcsec.

As discussed in Section 4.2.2, the Δt parameter controls the LBG duty cycle, weighting the abundance matching step towards haloes that have undergone a change in mass within the past Δt epoch. The motivation for this prescription is that such variation might correlate with recent bursts of star formation, and therefore UV luminosity and observability. Trenti et al. (2010) chose as their fiducial value $\Delta t = 200$ Myr, however we see that in our implementation this does not match the SILVERRUSH clustering signal. We also note that LAE selected galaxies tend to have younger ages than LBG selected populations (Gawiser et al., 2007).

We note again that the absence of any evolution in the observational data across these redshifts is puzzling. Even the $\Delta t = 20$ Myr duty cycle model does not achieve a low enough clustering to match the observations at $z = 6.6$.

We find that the luminosity function is insensitive to these tested variations in Δt.

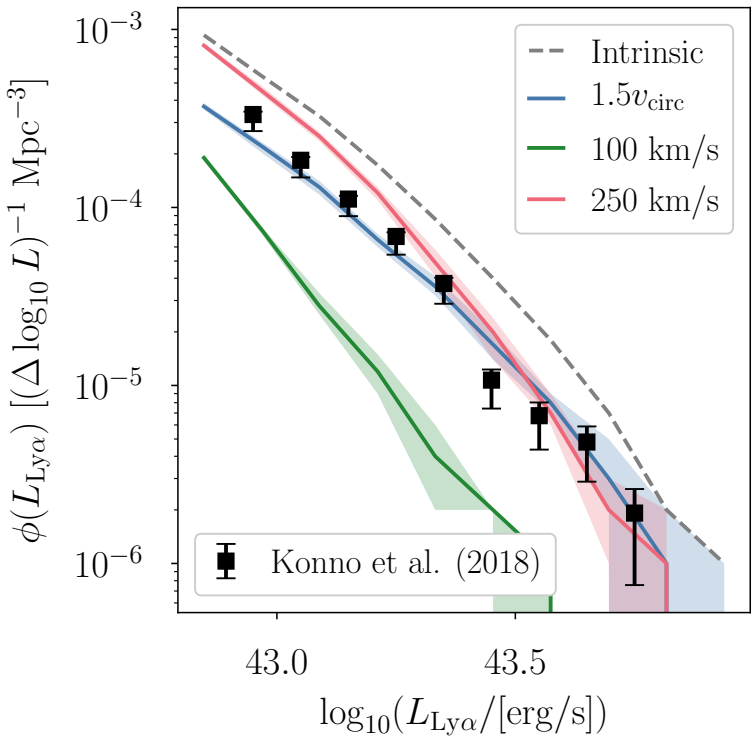


Fig. 4.13 Luminosity functions at $z = 5.756$, comparing different choices for the Δv emission parameter: our fiducial 1.5 v_{vcirc} in blue, a fixed value of 100 km/s in green, and a fixed value of 250 km/s in red. All lines are calculated using the same reionization history.

4.4.2 The effect of varying Δv on the luminosity function

In Figure 4.13 we show the luminosity function at $z = 5.756$, and the effect of varying the Δv parameter. As the transmission fraction is a strong function of the intrinsic emission profile (Dijkstra et al., 2011), the luminosity function is also dependent on our assumptions about this profile. We see that a fixed value of Δv does not fit well; for example the use of $\Delta v = 250$ km/s shown in red does fit the bright end reasonably well but overpredicts the number density of LAEs at the faint end. Fixing to a lower value, for example the fiducial choice of Chapter 3 of $\Delta v = 100$ km/s, results in too much attenuation by the IGM at all luminosities. Our fiducial choice in this work was to set $\Delta v \propto v_{\rm circ}$, which means that it scales with the LAE host halo mass as $\propto M_h^{1/3}$. Due to our population modelling (described in Section 4.2.2–4.2.4) the LAE luminosity should scale with the UV luminosity (with some stochasticity due to the REW distribution), and hence the host halo mass. This leads to the transmission distribution seen in Figure 4.10 and the good agreement in Figure 4.13.

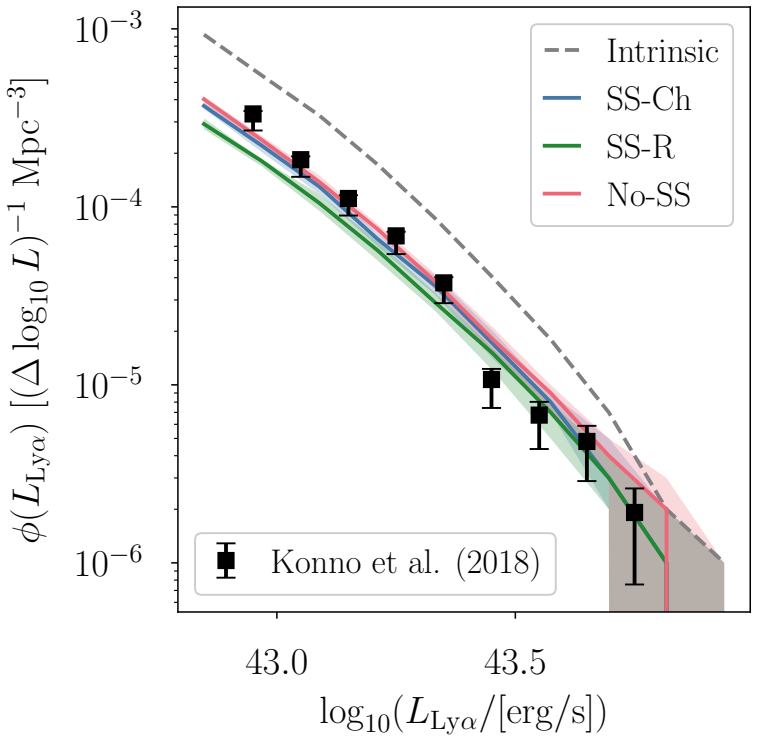


Fig. 4.14 Luminosity functions at $z = 5.756$, comparing different self-shielding prescriptions: our fiducial Chardin et al. (2018) based choice in blue, the prescription of Rahmati et al. (2013) in green, and also using no self-shielding prescription in red. All lines are calculated using the same reionization history.

4.4.3 The effect of self-shielding on the luminosity function

Finally in Figure 4.14 we show the luminosity function at $z = 5.756$, and the effect of varying the self-shielding prescription. Here we compare our fiducial self-shielding prescription with the cases

of no self-shielding or a Rahmati et al. (2013) prescription. We see that our luminosity function predictions are largely insensitive to the choice of prescription, although there is some increase in attenuation when self-shielding is added. We note that the non-negligible attenuation is present even in the absence of self-shielding; this is due to the non-zero neutral fraction that is found in photoionisation equilibrium at the outskirts of the halo.

4.5 Conclusions

In conclusion, in this chapter we have built an empirically constrained, self-consistent model of the evolution of LAEs in the epoch of reionization. This modelling made use of the halo population in the state-of-the-art Sherwood simulations, and the hydrodynamic gas structure for quantifying the transmission of Lyα emission through the IGM.

- We used the best fit REW probability distribution of Dijkstra & Wyithe (2012) to model the intrinsic REW distribution of our mock LAE population, and found that incorporating the IGM transmission fraction (calculated for each LAE individually) we reproduced the observed $z = 5.7$ REW distribution. Our transmission modelling assumes that the intrinsic LAE emission is a single Gaussian peak, with a velocity offset proportional to the host halo virial circular velocity. This gives a transmission fraction probability distribution across the mock population which corrects the overabundance of high equivalent widths predicted by the Dijkstra & Wyithe (2012) distribution at this redshift.

- In both our original reionization histories and the delayed-end versions, there is sufficient neutral hydrogen in the CGM and infalling gas further surrounding the halo at $z \sim 5.7$ that there is some attenuation redwards of Lyα. The attenuation by the neutral hydrogen in the CGM dominates (near the end of reionization) over the attenuation due to the large-scale IGM, even with the presence of residual neutral islands in the delayed-end models (when the average neutral fraction is non-zero). However at higher redshifts around the midpoint of reionization (when $\langle x_{\rm HII} \rangle \lesssim 0.5$) we find that variations in the neutral fraction of the wider IGM have a dominant effect on the Lyα transmission compared to the CGM neutral gas.

- Using this model we generated mock LAE populations at the redshifts of interest for narrowband surveys, and made predictions for the luminosity function and angular correlation function. Comparing these predictions with current data, in particular from the SILVERRUSH survey, we find that a rather late reionization history (our Very Late model) is in best agreement.

- In order to match the luminosity function across redshifts, we find that the delayed-end Very Late model has the best fitting evolution whilst still able to attenuate the signal enough at the highest redshifts.

- Employing a duty cycle in our LBG modelling allowed us to match the LAE 2-point correlation function with our mock population at $z = 5.7$. Our predictions for higher redshifts suggest that the ionisation structure of the IGM can enhance the clustering signal significantly already at $z = 6.6$. The lack of evolution in the current observed clustering at these redshifts is difficult to explain consistently with the evolution in the luminosity function and equivalent width distribution, and may suggest that the clustering at this redshift has not been measured with sufficient accuracy to extract the effect of reionization.

- In agreement with the results of Chapter 3, which found that comparing more and less massive host haloes there is a differential evolution in the relative transmission fraction (e.g. $T_{\rm IGM}(z)/T_{\rm IGM}(z = 5.7)$), we find that our transmission model also leads to a difference in the absolute transmission (e.g. $T_{\rm IGM}(z)$). We find that the more luminous LAEs are preferentially less attenuated by the IGM neutral fraction, albeit with a large scatter.

Lyman-α emitting galaxies have been considered as probes of reionization for over 20 years, and many attempts have been made at observing and modelling their behaviour at high redshifts. The ongoing ambitious Lyα surveys are starting to collect samples sufficiently large to allow us to put tight constraints on the reionization history of hydrogen. We find that the evolution of the luminosity function and angular correlation function are indeed strongly dependent on the reionization history, such that further observations at $z \geq 7$ and future Lyα surveys should allow us to map out in detail the second half ($\langle x_{\rm HII} \rangle \geq 0.5$) of reionization.

4.6 Additional Material

4.6.1 Comparing the use of full radiative transfer post-processing with the excursion set based method

In this section we compare three post-processing methods for constructing the large-scale ionisation field within the hydrodynamic simulation. We compare: (i) a simple excursion set prescription; (ii) the calibrated excursion set method used in Choudhury et al. (2015) (detailed in Section 2.2.4) and also in this work; and (iii) a full radiative transfer calculation. The full radiative transfer post-processing was performed using the Aton code (Aubert & Teyssier, 2008), as detailed in Kulkarni et al. (2019a). All methods were applied to the grids of the $L = 160$ cMpc/h, $N = 2 \times 2048^3$ Sherwood simulations to generate the ionised fraction in each grid cell. So that we are comparing like-for-like, we apply the same self-shielding prescription for all three methods to model the small-scale ionisation structure. We then generated mock LAE populations as detailed in Sections 4.2.2–4.2.4 and calculated the observable luminosity function and equivalent width distribution.

In Figure 4.15 we compare the luminosity functions predicted by these three methods. We show the Aton reionization model of Kulkarni et al. (2019a) in blue, the equivalent reionization history

implemented with the simple excursion set method in green, and finally our calibrated excursion set method in red. In order to compare like-for-like we have applied the excursion set method in the following two ways: firstly in the simple prescription we have used as inputs the mass-averaged ionised fraction, $\langle x_{\rm HII}\rangle_{\rm M}$, and the volume-averaged photoionisation rate within ionised regions, $\langle\Gamma_{\rm HI}\rangle_{\rm V}$, from the Aton fields. Given these two inputs we can apply the excursion set method and the self-shielding prescription to create the ionisation structure at large and small scales. Secondly we have used our calibrated method, in which we take the mass-averaged ionised fraction as before, but instead of using the Aton photoionisation rate directly we rather take the mean free path, $\lambda_{\rm mfp}$, from the Aton fields. This is used to solve for the background photoionisation rate consistently within our simulation volume (as detailed in Chapter 3, Section 2.2.4).

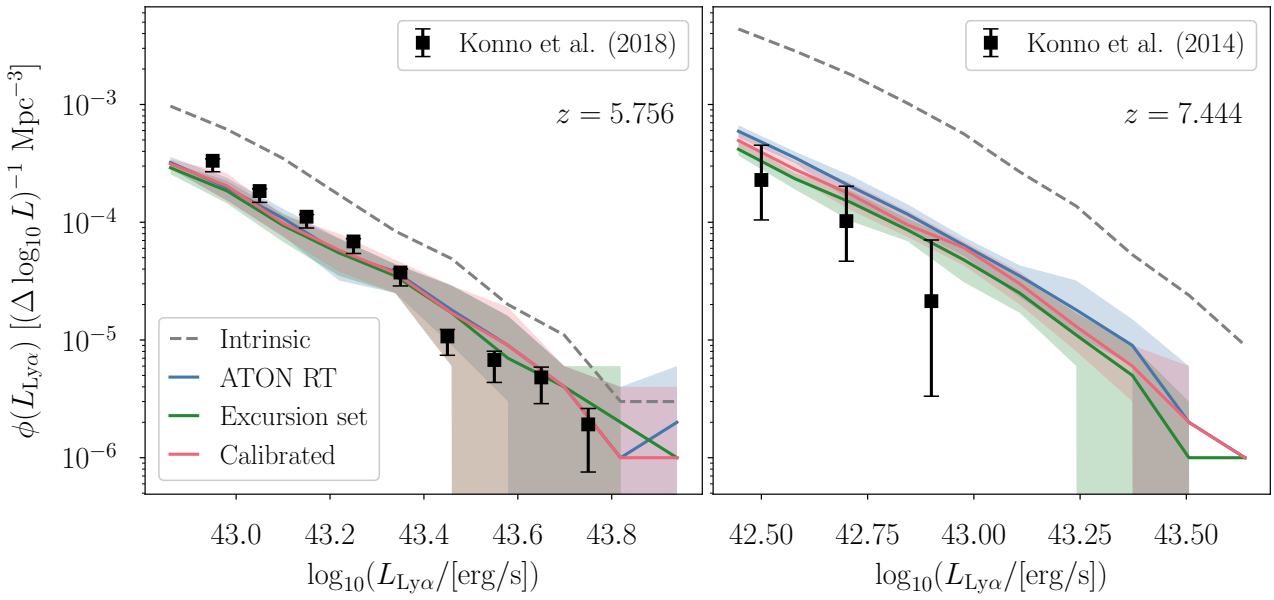


Fig. 4.15 Luminosity function at $z = 5.756$ and 7.444 for the same reionization history: calculated using full radiative transfer post-processing in blue, a simple excursion set implementation in green, and using our calibrated excursion set method in red.

We find that the three methods give similar predictions, consistent within the 68% scatter across the slices, and that the calibrated method is closer to the full Aton method at all redshifts. When the photoionisation rate is $\Gamma_{\rm HI} \gtrsim 10^{-13}{\rm s}^{-1}$ the simple excursion set model is close to the other models, however they start to diverge at higher redshifts when this is no longer the case. We find that the simple excursion set method, which assumes a uniform UV background, slightly overattenuates the luminosity function compared to the Aton method. We note that the full radiative transfer will not have a uniform UV background, and instead we would expect higher photoionisation rates near to the LAEs (where also the gas density is highest). This means that when we compare the neutral hydrogen densities around the LAEs, the Aton method gives a more ionised CGM compared to the excursion set method which sees the uniform UV background. This can be seen for $z = 7.444$ in Figure 4.15.

For the lower redshifts near to the end of reionization ($z \sim 6$) we find that the photoionisation rate is high enough that the ionised fraction saturates, such that the two methods are in good agreement, as seen in the left hand panel. In general our calibrated method predicts slightly higher background photoionisation rates which improves the agreement with the Aton results compared to the simple case.

4.6.2 Luminosity dependence of the CGM and IGM attenuation

In Section 4.3.2 we demonstrated that the CGM can play a significant role in the attenuation of Lyα, particularly relevant near the end of reionization. However in that section we considered only the median transmission of the full observed sample. We now quantify the variation in the role of CGM/IGM components across luminosity. We note that the Lyα luminosity of our sample broadly scales with the host halo mass (see Figure 4.11), hence the following also applies to the variation with mass. These results confirm what was found in Chapter 3.

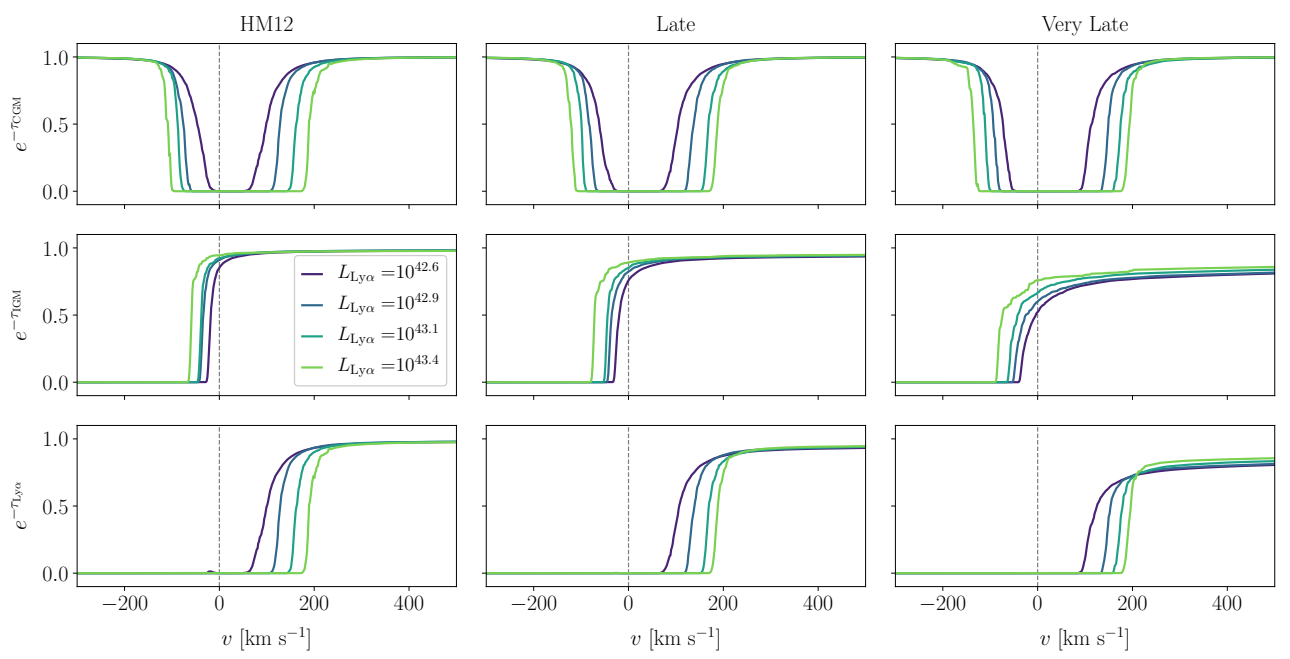


Fig. 4.16 The median transmission in different luminosity bins for the mock LAE population at $z = 7.444$, split into CGM (top), IGM (middle) and total (bottom) components as in Figure 4.6. The different (logarithmic) luminosity bins are indicated by the colour of the lines, with the centre of the bin indicated in the legend. From left to right we show the three original reionization histories: HM12, Late and Very Late.

In Figure 4.16 we show the median transmission curves, split into CGM, IGM and total components as in Figure 4.6, but calculated for samples in different luminosity bins. For brevity we show only the $z = 7.444$ original reionization history curves, from left to right showing the HM12, Late and Very

Late models respectively. The colour scale of the lines indicates the different luminosity bins, with the faintest bin centred on $10^{42.6}$ erg/s and the brightest centred on $10^{43.4}$ erg/s.

We see in Figure 4.16 that the attenuation from both the CGM and IGM components depends on luminosity. In particular for the CGM component we see that the fainter objects have more transmission around line-centre, compared to the brightest objects which have a wider absorption trough. This results from higher densities around the more massive haloes, which for the same UV background leads to more neutral gas around the brighter LAEs compared to the faint ones. In contrast for the IGM component, the brighter LAEs have more transmission compared to the faint ones. This is because the brighter LAEs (more massive haloes) reside in larger ionised regions compared to the fainter LAEs. As found in Kakiichi et al. (2016), these different luminosity dependences work against each other in the combined transmission.

Considering a given luminosity bin (one of the colours in Figure 4.16) we see that the CGM component is similar across the different reionization history models. In comparison the IGM component shows more difference, with the Very Late model in particular showing the most attenuation. This suggests that even with a CGM component that depends on the background photoionisation rate, for low enough average ionised fractions (at high enough redshifts into the second half of reionization) the LAE transmission is indeed a strong function of the IGM neutral fraction, and hence LAE observations can provide good reionization constraints.

We note finally that our modelling of the CGM makes various simplifying assumptions which may start to break down in the brightest LAEs. In particular we do not model the source emissivity when calculating the neutral hydrogen density around the halo. Furthermore with our assumption of an intrinsic emission profile we ignore the gas within $R_{\rm vir}$ of the halo centre. These limitations may affect the CGM transmission for the most massive (brightest) haloes.

4.6.3 Reionization history parameters

Our reionization prescription takes as input a reionization history given in terms of the mass-averaged ionised fraction evolution with redshift. However observers usually infer the volume-averaged fraction, which is weighted more towards volume-filling voids. In order to convert between the two quantities in practice is difficult; in the case of our simulations we can use a given snapshot (which is a realisation of the mass distribution) in order to measure both quantities.

Table 4.2 Volume-averaged neutral fractions, $\langle x_{\rm HI} \rangle_{\rm v}$, for the reionization histories considered in this thesis.

		Original			Delayed	
z	HM12	Late	Very Late	HM12	Late	Very Late
5.756	0.0000	0.0000	0.0000	0.0288	0.0377	0.0731
6.604	0.0000	0.1294	0.4168	0.1148	0.1654	0.4168
6.860	0.0395	0.2049	0.4849	0.1440	0.2132	0.4849
7.444	0.1734	0.3263	0.5999	0.2142	0.3263	0.5999

Table 4.3 Calibrated background photoionisation rates, $\log_{10}(\Gamma_{\rm HI}/{\rm s}^{-1})$, for the reionization histories considered in this thesis.

		Original			Delayed	
z	HM12	Late	Very Late	HM12	Late	Very Late
5.756	-12.50	-12.61	-12.64	-13.45	-13.47	-13.32
6.604	-12.79	-13.04	-13.26	-13.26	-13.24	-13.26
6.860	-12.88	-13.14	-13.26	-13.25	-13.21	-13.26
7.444	-13.07	-13.21	-13.27	-13.30	-13.20	-13.27

In Table 4.2 we show for each of our reionization history models the values of the volume-averaged neutral fraction, and similarly in Table 4.3 the photoionisation rates, at the redshifts we considered in this chapter. In Figure 4.17 we compare these values to a selection of observational constraints on the average neutral fraction of the IGM, on a logarithmic scale. These constraints were derived from observations of LAEs (Konno et al., 2018, 2014; Ouchi et al., 2018; Itoh et al., 2018), QSOs (Greig et al., 2017; Bañados et al., 2018) and GRBs (Totani et al., 2016, 2006; Greiner et al., 2009). We note that the points representing our reionization histories are not constraints but predictions of the models. The squares show the evolution in our original reionization histories, whilst the triangles show the delayed-end histories. Note that the delayed-end histories only deviate near the end of reionization, so for example the delayed-end Very Late model has the same neutral fraction evolution as the original for $z > 6$. The arrow indicates that because the original models have reionized by $z = 6$ (Late/Very Late) or $z = 6.7$ (HM12), the lower redshift datapoints where $\langle x_{\rm HI} \rangle_{\rm V} = 0$ are not visible within the logarithmic scale of the figure. We also show with a solid black curve the reionization history model of Kulkarni et al. (2019a) which was able to match opacity fluctuations in the Lyα forest, and note that it is very similar to our successful delayed-end Very Late model. In particular we highlight that both versions of our Very Late model are consistent with the SILVERRUSH observational constraints.

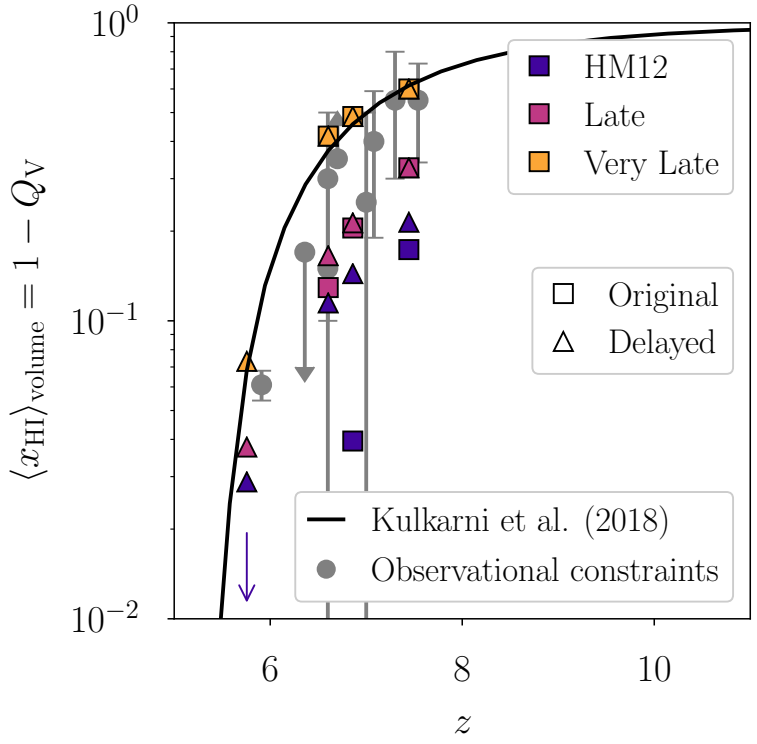


Fig. 4.17 The volume-averaged neutral fraction predicted in our models shown in squares (original histories) and triangles (delayed-end histories), compared to recent observational constraints from a variety of studies, shown with grey circles. These correspond to measurements of LAEs (Konno et al., 2018, $z = 6.6$), (Konno et al., 2014, $z = 7.3$), (Ouchi et al., 2018, $z = 6.6$), (Itoh et al., 2018, $z = 7$); QSOs (Greig et al., 2017, $z = 7.1$), (Bañados et al., 2018, $z = 7.5$); and GRBs (Totani et al., 2016, $z = 5.9$), (Totani et al., 2006, $z = 6.4$), (Greiner et al., 2009, $z = 6.7$). The black solid line shows the model of Kulkarni et al. (2019a), which is similar to our delayed-end Very Late model, and was found to reproduce the opacity fluctuations in the Lyα forest. The vertical scale of this figure is logarithmic.

Chapter 5

Probing delayed-end reionization histories with the 21cm-LAE cross-power spectrum

In this chapter we model the 21-cm signal and LAE population evolution during the epoch of reionization in order to predict the 21cm-LAE cross-power spectrum. We employ high-dynamic-range simulations of the IGM to create models that are consistent with constraints from the CMB, Lyman-α forest and LAE population statistics. Using these models we consider the evolution of the cross-power spectrum for a selection of realistic reionization histories and predict the sensitivity of current and upcoming surveys to measuring this signal. We find that the imprint of a delayed end to reionization can be observed by future surveys, and that strong constraints can be placed on the progression of reionization as late as $z = 5.7$ using a Subaru-SKA survey. We make predictions for the signal-to-noise ratios achievable by combinations of Subaru/PFS with the MWA, LOFAR, HERA and SKA interferometers for an integration time of 1000 hours. We find that a Subaru-SKA survey could measure the cross-power spectrum for a late reionization at $z = 6.6$ with a total signal-to-noise greater than 5, making it possible to constrain both the timing and bubble size at the end of reionization. Furthermore, we find that expanding the current Subaru/PFS survey area and depth by a factor of three would double the total signal-to-noise.

Declaration

The work presented in this chapter was published as a paper titled "*Probing delayed-end reionization histories with the 21cm-LAE cross-power spectrum*", published in Monthly Notices of the Royal Astronomical Society, Volume 494, Issue 1, p. 703-718. This work was done in collaboration with Girish Kulkarni and Martin G. Haehnelt. The underlying Sherwood simulations that form the basis for the modelling in this work were performed by a prior collaboration (Bolton et al., 2017). Otherwise the modelling and analysis was entirely performed by this author, with guidance provided by the collaborators.

5.1 Background

During the cosmic period of history known as the epoch of reionization (EoR) the first stars and galaxies altered the large-scale ionisation state of the Universe, with consequences persisting down to the present epoch (Dayal & Ferrara, 2018). In particular this period represents a phase transition in which the primordial neutral hydrogen in the intergalactic medium (IGM) — left over from recombination (Barkana & Loeb, 2001) — was ionised by sources such as galaxies and possibly early quasars (Loeb & Barkana, 2001). Observations of the IGM in the high redshift Universe (such as the Lyman-α forest seen in sightlines to z = 5–6 quasars, McGreer et al., 2015) constrain the neutral hydrogen fraction to be $\langle x_{\rm HI} \rangle < 0.1$ by $z \sim 6$. The reionization process involved radiative transfer across a range of scales, from within galaxies (Katz et al., 2019b) to the farthest reaches of the IGM (Miralda-Escudé et al., 2000). Its impacts included photoheating of the IGM (Miralda-Escudé & Rees, 1994) and the inhibition of star formation in dwarf galaxies (Efstathiou, 1992; Katz et al., 2019a). From a theoretical standpoint, reionization can indirectly give us constraints on the first populations of stars and early galaxies (Duncan & Conselice, 2015).

A number of methods have been established to measure reionization. One direct probe of the ionisation state of the IGM is through the 21-cm signal from neutral hydrogen (Pritchard & Loeb, 2012; Mesinger et al., 2011; Furlanetto et al., 2006). Radiation at a rest-frame wavelength of 21 cm is emitted and absorbed by neutral hydrogen via the hyperfine transition of the ground state (Wild, 1952). It should be possible to observe the 21-cm signal from the high redshift IGM, relative to the cosmic microwave background (CMB) radiation, using radio interferometry (Tozzi et al., 2000). Observations of the 21-cm signal are complicated by foreground radio signals (Sims et al., 2016; Chapman et al., 2015) which must be removed or avoided (Liu et al., 2014).

Another indirect probe of reionization is the population statistics of Lyman-α emitters (LAEs), such as the luminosity and angular clustering functions that we considered in Chapter 4. Intervening neutral hydrogen gas can obscure the Lyman-α emission from distant LAEs, so we expect to see an attenuation of the luminosity function at redshifts within the EoR (Gunn & Peterson, 1965; Madau & Rees, 2000; Dijkstra et al., 2007). Interpreting LAE observations in the context of reionization is challenging because the radiative transfer processes occurring within the galaxy's ISM — including the interaction with dust — is degenerate with attenuation by the IGM (Hutter et al., 2014, 2015). Alongside the attenuating effect on the luminosity function, another imprint of reionization on the high-z LAE population is the enhancement of angular clustering. We expect LAEs to reside within the overdensities that sit inside the first ionised bubbles, and this spatial correlation can enhance the observed clustering signal (McQuinn et al., 2007b).

There has been extensive theoretical work modelling both the 21-cm signal (including Beane & Lidz, 2018; Shimabukuro & Semelin, 2017; Kulkarni et al., 2017; Hassan et al., 2016; Kulkarni et al., 2016; Mesinger et al., 2011; McQuinn et al., 2006) and the LAE population (including Laursen et al., 2019; Inoue et al., 2018; Weinberger et al., 2019; Mason et al., 2018b; Kakiichi et al., 2016; Hutter et al., 2015; Gronke et al., 2015; Hutter et al., 2014; Zheng & Wallace, 2014; Jensen et al., 2013; Dijkstra

et al., 2007) during the epoch of reionization. The epoch of reionization has been explored with analytic prescriptions (such as Choudhury, 2009; Furlanetto et al., 2004c), using semi-numerical modelling (such as Mesinger et al., 2011; Majumdar et al., 2014), and full numerical simulation (such as Trac & Cen, 2007; Rosdahl et al., 2018; Chardin et al., 2015). The challenge of modelling reionization is to try to capture the physical processes which occur on extremely different scales: from the large-scale ionised bubble structure (McQuinn et al., 2007a; Zahn et al., 2007; Giri et al., 2018a), to the small-scale self-shielding of neutral gas clumps (Rahmati et al., 2013; Chardin et al., 2018). Accurately modelling this large dynamic range is particularly important for the 21-cm signal (Kulkarni et al., 2016). Similarly predicting the effect of reionization on the observability of LAEs is complicated by resonant radiative transfer of the Lyman-α emission within the galaxy and the wider halo environment (Sadoun et al., 2017; Dijkstra, 2014). In this chapter we employ the 21-cm model used in Kulkarni et al. (2016), combined with the reionization histories and LAE modelling described in Chapter 4.

The cross-correlation of the 21-cm signal and the LAE population at high redshifts has been considered previously by a number of authors. Early work considered the possibility of measuring the cross-power spectrum (Furlanetto & Lidz, 2007; Lidz et al., 2009), predicting observational sensitivities for instruments such as the Murchison Widefield Array (MWA, Park et al., 2014), the LOW Frequency ARray (LOFAR, Wiersma et al., 2013; Vrbanec et al., 2016) and the Square Kilometer Array (SKA, Sobacchi et al., 2016). The cross-correlation of these two observables is not subject to the radio foregrounds that plague the interpretation of individual 21-cm observations. Many theoretical approaches use computational simulations of the IGM to model the cross-correlation signal, but it has also been considered analytically (for example, in Feng et al., 2017). Recent work has explored the synergy of the SKA and Subaru, considering both the power spectrum and the correlation function. These studies found that SILVERRUSH type Subaru surveys combined with the SKA are optimal for distinguishing an IGM with average neutral fractions of 50%, 25% and 10% at $z = 6.6$ (Hutter et al., 2018, 2017). These works also established that deeper and wider LAE surveys can improve the sensitivity to measuring the cross-power spectrum (Kubota et al., 2019; Yoshiura et al., 2018; Kubota et al., 2018). Another interesting area of research is the cross-correlation of the 21-cm signal with other types of galaxies, for example [O III] emitters (Moriwaki et al., 2019), or with intensity maps of other emission lines (Neben et al., 2017).

In this chapter we employ our empirical models with the high-dynamic-range Sherwood simulations to test the effect of six reionization histories — defined in Section 4.2 of the previous chapter — on the cross-correlation evolution, including the delayed-end history that has successfully explained the large opacity fluctuations in the Lyα forest (Kulkarni et al., 2019a; Keating et al., 2020; Nasir & D'Aloisio, 2019). We make predictions for a selection of observational surveys using updated parameters, but following the framework established by earlier work (Furlanetto & Lidz, 2007; Lidz et al., 2009, in particular). Our aim is to determine whether observations of 21cm-LAE cross-correlations can provide additional evidence to support a delayed end to reionization. In order to determine this, we explore how sensitive the different surveys are to measuring the 21cm-LAE cross-power spectrum

in a delayed reionization scenario, and also to what extent this scenario can be distinguished from earlier reionization histories.

This chapter is structured as follows: in Section 5.2 we detail the setup of our simulations and theoretical models; in Section 5.3 we present our results for the cross-power spectrum and our predictions for observational sensitivities; in Section 5.4 we discuss the implications of these results; and finally we conclude in Section 5.5. In particular, Section 5.2 is split into subsections describing: our simulations (§5.2.1), our model for the 21-cm signal (§5.2.2), our LAE population model (§5.2.3), and the definitions we use for the power spectra calculations (§5.2.4). Alongside this, Section 5.6 provides additional material including: a more detailed description of the numerical implementation of our power spectrum calculation (§5.6.1); derivations of our sensitivity modelling (§5.6.2); a presentation of the cross-correlation functions derived from our cross-power spectra results (§5.6.3); and finally a further discussion of the impact of our self-shielding assumptions (§5.6.4).

5.2 Method

5.2.1 Simulating reionization

In this sub-section we summarise the key features of our simulation set-up, but we note that it is the same as that used in Chapters 3 & 4, so refer the reader to those chapters for further details.

Cosmological simulations

The basis for our theoretical modelling is the Sherwood suite of cosmological hydrodynamic simulations (Bolton et al., 2017), designed to achieve a high-dynamic-range and effectively capture the behaviour of the low-density IGM. As in Chapter 4 we make use of a periodic box of side $L = 320$ cMpc/h, run using the P-Gadget-3 (Springel, 2005; Springel et al., 2001) SPH code with $N = 2 \times 2048^3$ particles ($M_{\rm DM} = 2.75 \times 10^8$ h$^{-1}M_\odot$). A Friends-of-Friends algorithm was used to identify dark matter haloes on-the-fly. This simulation used the Planck Collaboration et al. (2014) cosmological parameters: $h = 0.678$, $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_b = 0.0482$, $\sigma_8 = 0.829$, $n = 0.961$, and $Y_{\rm He} = 0.24$. When calculating the 21-cm brightness temperature, the SPH kernel was used to interpolate particles onto a uniform grid.

The Sherwood suite has been used to model a variety of phenomena such as the 21-cm reionization signal (Kulkarni et al., 2016, 2017), the LAE evolution at high redshift (Weinberger et al., 2018, 2019), the detectability of Ly-α emission in the cosmic web (Witstok et al., 2019), a late reionization and the opacity fluctuations in the Ly-α forest (Keating et al., 2020; Kulkarni et al., 2019a), and the detectability of [C II] line intensity mapping (Dumitru et al., 2019).

Modelling the IGM ionisation structure

In order to explore the effect of reionization on the 21cm-LAE cross-power spectrum we implement the calibrated reionization modelling of Choudhury et al. (2015) outlined in Section 2.2. This involves first generating the large-scale ionisation structure using an excursion set approach (Furlanetto et al., 2004c; Mesinger & Furlanetto, 2007; Choudhury et al., 2009; Mesinger et al., 2011; Santos et al., 2010; Hassan et al., 2016), followed by a calculation of the ionisation state of the small-scale ionised regions which are assumed to be in photoionisation equilibrium. For a given reionization history which maps the average ionised fraction of the IGM, $\langle x_{\rm HII}(z)\rangle$, the two steps of large- and small-scale ionisation structure are calibrated to find a consistent UV-background photoionisation rate, $\Gamma_{\rm HI}(z)$. This allows us to impose arbitrary reionization histories specified by $\langle x_{\rm HII}(z)\rangle$, and calculate the resulting ionisation structure self-consistently across the scales probed by the simulation.

As part of the photoionisation equilibrium calculation, we employ the self-shielding prescription of Chardin et al. (2018) described in Section 2.2.5. This is a redshift-dependent modification of the Rahmati et al. (2013) parametrisation, which accounts for changes to the local photoionisation rate in the proximity of overdense self-shielded gas.

Reionization histories

We test six physically motivated reionization histories in this chapter, using the same models as in Chapter 4. These histories are based on the original three reionization histories explored by Choudhury et al. (2015):

- **HM12**, based on the minimal reionization model of Haardt & Madau (2012). This is the earliest reionization model considered in this work, with reionization ending by $z = 6.7$.
- **Late**, similar to the HM12 evolution but shifted in redshift so that it ends at a later time of $z = 6$.
- **Very Late**, a further variation of the Late evolution, in which the end of reionization remains at $z = 6$ but the gradient $\mathrm{d}\langle x_{\rm HII}\rangle/\mathrm{d}z$ is altered to be more abrupt.

We use these three models (hereafter referred to as the *original* models), as well as three modified versions (referred to as the *delayed-end* models) in which the end of reionization is delayed to $z \sim 5.3$.

In particular we note that the delayed-end Very Late model has a very similar evolution to the reionization histories considered in Kulkarni et al. (2019a) & Keating et al. (2020), which successfully simulated the opacity fluctuations seen in observations of the Ly-α forest (Becker et al., 2015b). In Chapter 4, the LAE evolution predicted by this delayed-end Very Late model was found to be consistent with observations of the equivalent width distribution, luminosity function and angular correlation function as measured by the SILVERRUSH survey (Konno et al., 2018; Shibuya et al., 2018; Itoh et al., 2018). We also note that the lower value of $\tau = 0.054 \pm 0.007$ measured by Planck Collaboration (2018b) is consistent with a late end to reionization. This reionization model therefore

offers a neutral fraction evolution that is consistent with many independent observational constraints on reionization.

The average ionised fraction evolution for each of our six reionization histories can be seen in Figure 5.1.

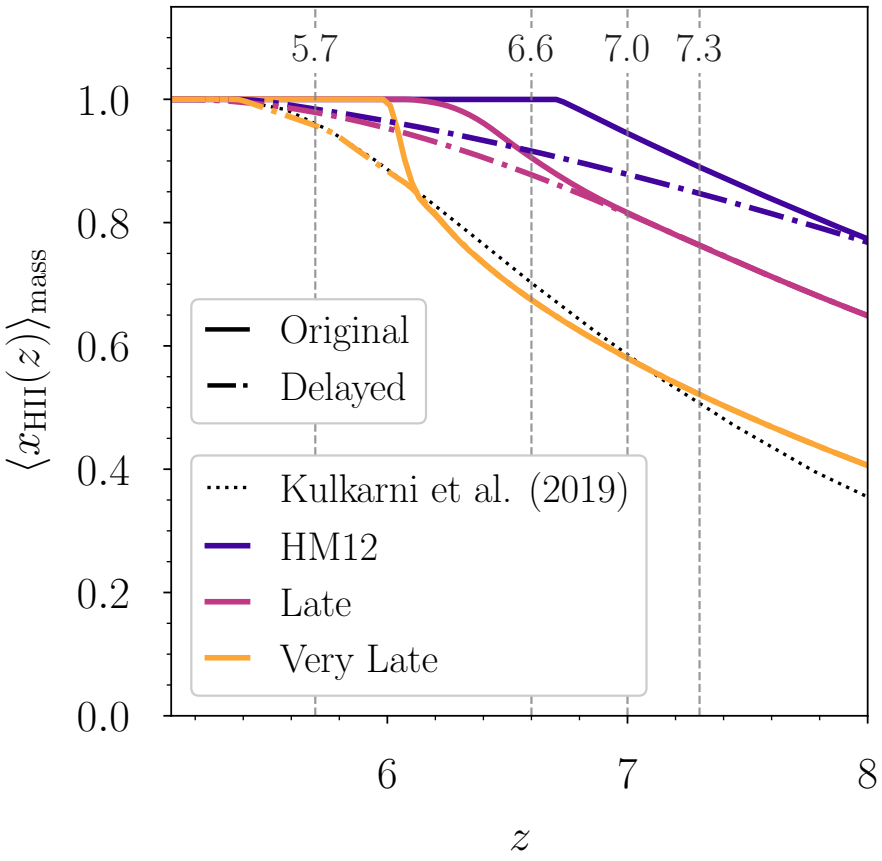


Fig. 5.1 The mass-averaged ionised fraction for the six reionization histories considered in this work, showing the *original* models with solid lines and the *delayed-end* models with dot-dashed lines. The positions of the Subaru narrow-band filter centres are shown with the grey dashed lines. For comparison the reionization history considered in Kulkarni et al. (2019a) is shown with the black dotted line.

5.2.2 Simulating 21-cm emission

We follow the approach of Kulkarni et al. (2016) to model the 21-cm brightness temperature using the Sherwood simulations. We assume that the spin temperature of hydrogen is much greater than the cosmic microwave background (CMB) temperature, which should be valid for the redshifts considered in this work where the average ionised fraction of the IGM is greater than a few percent (Pritchard & Loeb, 2012). We can then calculate the brightness temperature of the 21-cm line relative to the CMB

as a function of the H I fraction, $x_{\rm HI}$, and density contrast, $\Delta = \rho/\bar{\rho}$, using (Furlanetto et al., 2006),

$$T_{21}(\mathbf{x}) = \overline{T}_{21}\, x_{\rm HI}(\mathbf{x})\, \Delta(\mathbf{x}), \tag{5.1}$$

where the brightness temperature of the gas at mean density is (Choudhury et al., 2009),

$$\overline{T}_{21} \simeq 22\,{\rm mK}\,\left[(1+z)/7\right]. \tag{5.2}$$

In order to calculate this for our simulations we use the SPH kernel to interpolate the SPH particles onto a uniform grid with 1024 cells on a side (cell size $L_{\rm cell} = 312.5$ ckpc/h) allowing us to calculate the density contrast at a given cell position, $\mathbf{x}$. We apply the reionization history modelling as described in section 5.2.1 to determine the H I fraction in each grid cell.

5.2.3 Simulating Lyman-α emitters

We employ the LAE model described in Chapter 4. As in section 5.2.1 we will now summarise the key features of this model, and refer the reader to the previous chapter for further details. An LAE population is generated within the simulation in the following stages:

- **Identify LBG galaxies**: The first stage of the model is to determine which dark matter haloes host LBG-like galaxies. We abundance match the dark matter halo mass function with observed UV luminosity functions at $z = 5.9$, using a duty-cycle following the prescription of Trenti et al. (2010). This duty cycle is parametrised by Δt, which gives a measure of the time period over which the observability of the LBG population is evolving. This creates a mapping between halo mass, M_h, and LBG UV luminosity, $L_{\rm UV}$. We assume this mapping applies to other redshifts, or in other words that the LBG population evolution is driven by the underlying halo population evolution (at high redshifts). This means we can apply the mapping at other redshifts to derive a population of LBGs (a subset of the halo population) with assigned UV luminosity, $L_{\rm UV}$.

- **Identify LAE galaxies**: The second stage of the model identifies which of the LBG population has sufficient Lyman-α emission to be observed as an LAE. We employ the Lyman-α rest-frame equivalent width (REW) distribution model of Dijkstra & Wyithe (2012), which assigns a conditional probability for an LBG having a certain Lyman-α REW given its UV magnitude, $P({\rm REW}|M_{\rm UV})$. In particular we assume this model describes the *intrinsic* REW distribution of the LBG population, when attenuation of Lyman-α radiation by the IGM is neglected. For each LBG we sample from this distribution to derive its intrinsic REW (and hence also its intrinsic Lyman-α luminosity), thus generating an LAE population.

- **Apply observational selection**: The final stage of the model is to determine which LAEs would be observed, including the attenuation of their emission by the IGM, given the selection function of a particular survey. We employ observational limits on the LAE REW and $L_{{\rm Ly}\alpha}$

based on the SILVERRUSH surveys, shown in Table 5.1. To calculate the radiative transfer of Lyman-α emission from the LAEs to the observer, we extract sightlines through the simulation (with ionisation structure derived as in section 5.2.1). We assume the complex radiative transfer within the halo produces a double-peaked emission profile, of which only the red peak will escape through the IGM. We model this intrinsic emission profile as a Gaussian peak (Choudhury et al., 2015), with a velocity offset that scales with the host halo virial velocity $\Delta v \propto v_{\rm circ}$. For the radiative transfer throughout the rest of the IGM, we calculate the effect of scattering out of the line-of-sight using the $e^{-\tau}$ approximation (Laursen et al., 2011). We then calculate the Lyman-α transmission fraction, $T^{\rm IGM}_{{\rm Ly}\alpha}$, along a single sightline to each LAE so that we can finally determine their observed luminosity, $L^{\rm obs}_{{\rm Ly}\alpha} = T^{\rm IGM}_{{\rm Ly}\alpha} L^{\rm int}_{{\rm Ly}\alpha}$. Note we are therefore neglecting any intrinsic or dust-driven evolution of the emission profile.

We apply our LAE model to four snapshots of our Sherwood simulation, at redshifts $z = 5.756$, 6.604, 6.860, 7.444 which are close to the central redshifts measured by the NB816, NB921, NB973, NB101 filters on the Hyper Suprime-Cam (HSC) of the Subaru telescope. This generates mock LAE populations which could be observed with narrowband observations using the Subaru telescope.

Table 5.1 Observational selection thresholds (as in Chapter 4) used to generate mock observed samples.

Based on survey	z	REW$_{\rm min}$ [Å]	$L_{{\rm Ly}\alpha,{\rm min}}$ [erg/s]
Konno et al. (2018)	5.7	10	6.3×10^{42}
Konno et al. (2018)	6.6	14	7.9×10^{42}
Ota et al. (2017) Itoh et al. (2018)	7.0	10	2×10^{42}
Konno et al. (2014)	7.3	0	2.4×10^{42}

Importantly the key free parameters in the LAE model are the duty cycle parameter, Δt, and the velocity offset, Δv. We calibrate these parameters at $z = 5.7$, such that our mock population matches the observed REW distribution, luminosity function and angular correlation function as measured by Shibuya et al. (2018); Konno et al. (2018); Ouchi et al. (2018). These calibrated parameters are assumed to be fixed for the other redshifts, such that any evolution in the LAE population is due to the underlying halo population and IGM neutral fraction evolution. The velocity offset, $\Delta v = a\, v_{\rm circ}$, is calibrated to,

$$a = \begin{cases} 1.5 & \text{Original,} \\ 1.8 & \text{Delayed,} \end{cases} \tag{5.3}$$

for the two different reionization history cases. In Chapter 4 it was found that — after calibrating at $z = 5.7$ — these models compare favourably to the observed LAE populations in the SILVERRUSH survey, in particular predicting the evolution of the luminosity function, equivalent width distribution and angular correlation function seen at the other narrowband redshifts.

In Figure 5.2 we plot a slice of our simulation volume, showing a projection of the 21-cm brightness temperature and the positions of LAEs. We note that the LAEs are predominately found in the overdense regions which ionise first, and hence we mostly see LAEs in regions where the 21-cm signal is faint. Conversely the underdense regions which remain neutral for longer do not host many LAEs, and so we see a brighter 21-cm signal where there are fewer LAEs.

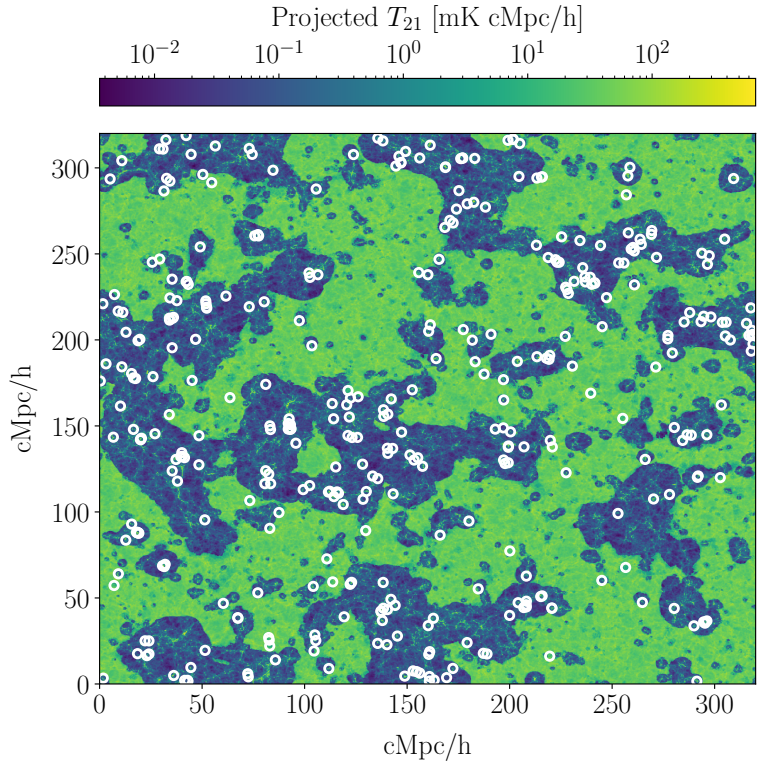


Fig. 5.2 The 21-cm brightness temperature projected over a thin 2.5 cMpc/h slice of the simulation at $z = 6.604$ for the delayed-end Very Late model. Overplotted with white empty circles are the positions of simulated LAEs within this thin slice.

5.2.4 Calculating the power spectrum

We work with the brightness temperature $T_{21\mathrm{cm}}(\mathbf{x})$ directly, and the LAE number density contrast defined as,

$$\delta_{\mathrm{LAE}} \equiv \frac{n_{\mathrm{LAE}}(\mathbf{x}) - \langle n_{\mathrm{LAE}} \rangle}{\langle n_{\mathrm{LAE}} \rangle}, \tag{5.4}$$

where $\langle . \rangle$ indicates the spatial average. As described above, the 21-cm brightness temperature, $T_{21\mathrm{cm}}$, is modelled using our simulations on a uniform grid with 1024^3 cells (with cell size 0.3125 cMpc/h). We interpolate the LAE population onto the same grid geometry using the Cloud-in-Cell (CIC) scheme. We can then calculate the cross-power spectrum as,

$$\langle \tilde{T}_{21\mathrm{cm}}(\mathbf{k_1}) \tilde{\delta}_{\mathrm{LAE}}(\mathbf{k_2}) \rangle = (2\pi)^3 \delta_D(\mathbf{k_1} + \mathbf{k_2}) P_{21\mathrm{cm}\times\mathrm{LAE}}(k), \tag{5.5}$$

where the tilde quantity $\tilde{X}(\mathbf{k})$ denotes the Fourier transform of the field $X(\mathbf{x})$, δ_D is the Dirac delta function and $P(k)$ is the power spectrum which depends only on $k = |\mathbf{k}|$ as a result of the statistical isotropy of the fields. Similarly the auto-power spectra are calculated using,

$$\langle \tilde{T}_{21\mathrm{cm}}(\mathbf{k_1})\tilde{T}_{21\mathrm{cm}}(\mathbf{k_2})\rangle = (2\pi)^3\delta_D(\mathbf{k_1}+\mathbf{k_2})P_{21\mathrm{cm}}(k), \tag{5.6}$$

$$\langle \tilde{\delta}_{\mathrm{LAE}}(\mathbf{k_1})\tilde{\delta}_{\mathrm{LAE}}(\mathbf{k_2})\rangle = (2\pi)^3\delta_D(\mathbf{k_1}+\mathbf{k_2})P_{\mathrm{LAE}}(k). \tag{5.7}$$

Further details of the numerical implementation of this calculation can be found in the additional materials in Section 5.6.1.

When presenting our results we will hereafter work with dimensionless power spectra defined as,

$$\Delta^2(k) \equiv \frac{k^3}{2\pi^2}P(k). \tag{5.8}$$

This quantity carries no length dimensionality, however given the conventions defined in Eqs. (5.5) & (5.6) we note that $\Delta^2_{21\mathrm{cm}}(k)$ has units of temperature squared, $(\mathrm{mK})^2$, whilst $\Delta^2_{21\mathrm{cm}\times\mathrm{LAE}}(k)$ has units of temperature, mK. We can also calculate the cross-correlation coefficient,

$$r_{21\mathrm{cm}\times\mathrm{LAE}}(k) = \frac{P_{21\mathrm{cm}\times\mathrm{LAE}}(k)}{\sqrt{P_{21\mathrm{cm}}(k)P_{\mathrm{LAE}}(k)}}. \tag{5.9}$$

5.2.5 Calculating observational sensitivities

We follow the approach of Furlanetto & Lidz (2007) & Lidz et al. (2009) for exploring the errors introduced in the measurements of the 21-cm signal and the LAE survey. For a particular $\mathbf{k}$-mode measured along a given line-of-sight labelled by μ (the cosine of the angle between the line-of-sight and $\mathbf{k}$), and restricting ourselves to the upper-half $\mathbf{k}$-plane, the variance on the cross-power spectrum is given by (Dumitru et al., 2019),

$$\mathrm{var}[P_{21\mathrm{cm}\times\mathrm{LAE}}(k,\mu)] = \frac{1}{2}\left[P^2_{21\mathrm{cm}\times\mathrm{LAE}}(k,\mu) + \sqrt{\mathrm{var}[P_{21\mathrm{cm}}(k,\mu)]\,\mathrm{var}[P_{\mathrm{LAE}}(k,\mu)]}\right], \tag{5.10}$$

where the individual survey variances are given by,

$$\mathrm{var}[P_{21\mathrm{cm}}(k,\mu)] = \left[P_{21\mathrm{cm}}(k,\mu) + P^{\mathrm{noise}}_{21\mathrm{cm}}(k,\mu)\right]^2, \tag{5.11}$$

and,

$$\mathrm{var}[P_{\mathrm{LAE}}(k,\mu)] = \left[P_{\mathrm{LAE}}(k,\mu) + P^{\mathrm{noise}}_{\mathrm{LAE}}(k,\mu)\right]^2, \tag{5.12}$$

Full derivations for the noise terms, $P^{\mathrm{noise}}(k,\mu)$, can be found in the additional materials in Section 5.6.2. When binned in both k and μ, an annulus of width $(\Delta k, \Delta\mu)$ will contain a number of Fourier modes,

N_m, given by

$$N_m(k,\mu) = 2\pi k^2 \Delta k \Delta\mu \left(\frac{V_{\rm survey}}{(2\pi)^3}\right), \tag{5.13}$$

where $V_{\rm survey}$ is the comoving survey volume (see Section 5.6.2 for further discussion of volume matching). In practice we are interested in the variance of the spherically-averaged power spectrum, which can be obtained by summing the errors over μ in inverse quadrature. Hence we can estimate the variance on a given k-mode of the spherically-averaged power spectrum as,

$$\frac{1}{{\rm var}[P_{\rm 21cm\times LAE}(k)]} = \sum_{\mu} \frac{N_m(k,\mu)}{{\rm var}[P_{\rm 21cm\times LAE}(k,\mu)]}. \tag{5.14}$$

For a given power spectrum we estimate the line-of-sight enhancement due to redshift-space distortions as,

$$P(k,\mu) = (1+\beta\mu^2)^2 P(k). \tag{5.15}$$

Here $\beta = \Omega_{\rm m}^{0.6}(z)/b$, where b is the bias factor (Kaiser, 1987) of the respective field. We assume $b_{\rm 21cm} = 1$ and calculate,

$$b_{\rm LAE}^2 = P_{\rm LAE}(k)/P_{\rm DM}(k), \tag{5.16}$$

where $P_{\rm DM}(k)$ is the dark matter power spectrum. We employ this linear treatment to avoid the computational overhead of the full radiative transfer calculation (i.e. using the simulation gas peculiar velocities), although note that algorithms for the full treatment do exist (such as Chapman & Santos, 2019).

In this work we consider 21-cm observations using the Phase II MWA (Wayth et al., 2018), LOFAR (van Haarlem et al., 2013), HERA (DeBoer et al., 2017), and SKA1-low (Waterson et al., 2016) interferometers. For the LAE galaxy survey we estimate the sensitivities of a Subaru survey, using a combination of narrowband selection with the Hyper Suprime-Cam (HSC) (Miyazaki et al., 2018; Ouchi et al., 2018) and spectroscopy from the Prime Focus Spectrograph (PFS) (Tamura et al., 2018; Takada et al., 2014). We assume an LAE survey with the field of view of the HSC Subaru Strategic Program's (SSP) Deep field (which has an area of 27 deg^2, Miyazaki et al., 2018) when calculating LAE sensitivities. Details of the sensitivity modelling of these instruments can be found in the additional materials in Section 5.6.2, whilst the chosen parameters are shown in Table 5.2.

Table 5.2 Survey parameters for modelling observational sensitivities, as detailed in Section 5.6.2. For all the 21-cm surveys we assume an integration time of $t = 1000$ hours, a bandpass of $B = 8$ MHz and a spectral resolution of $\Delta\nu = 50$ kHz, hence the available parallel k-modes at $z = 6.6$ lie in the interval between $k_{\parallel,\min} = 0.056$ cMpc^{-1} and $k_{\parallel,\max} = 8.983$ cMpc^{-1}.

Parameters	MWA[a]	LOFAR[b]	HERA[c]	SKA1[d]	Subaru[e]
Effective collecting area, $A_e^{\dagger}$ [m^2]	21.5	512.0	77.6	497.4	–
Number of antennae, N_a	256‡	24	350	512	–
Core radius, $r_{\rm core}$ [m]	100	160	150	350	–
Maximum radius, $r_{\max}$ [km]	3.5	2.0	0.45	6.4	–
Minimum baseline, $b_{\min}$ [m]	7.7	68.0	14.6	35.0	–
Maximum baseline, $b_{\max}$ [km]	5.3	3.5	0.876	12.8	–
Minimum transverse k-mode at $z = 6.6$, $k_{\perp,\min}$ [cMpc^{-1}]	0.003	0.031	0.007	0.016	–
Maximum transverse k-mode at $z = 6.6$, $k_{\perp,\max}$ [cMpc^{-1}]	2.405	1.588	0.397	31.761	–
Spectroscopic redshift error, σ_z	–	–	–	–	0.0007
Survey volume at $z = 6.6$, $V_{\rm survey}$ [cMpc^3]	1.6×10^9	6.6×10^7	4.3×10^8	6.8×10^7	2.3×10^7
Field-of-view, $\Omega^{\dagger}$ [deg^2]	610	26	169	26	27

[a] Extended configuration, based on Wayth et al. (2018); Tingay et al. (2013).
[b] NL-Core configuration, based on van Haarlem et al. (2013).
[c] Hera-350 configuration, based on DeBoer et al. (2017).
[d] SKA1-low V4A core configuration, based on SKAO Science Team (2015); Chang et al. (2015).
[e] We assume a narrowband ($\Delta z = 0.1$) survey geometry, based on Miyazaki et al. (2018); Takada et al. (2014).
† Observing at 150 MHz. We assume a frequency dependence of $A_e(\nu)/A_e(\nu = 150\text{MHz}) = (150/\nu)^2$.
‡ The Phase II upgrade gives the MWA a total of 256 tiles, however with the existing receivers and correlator it is only possible to use 128 at any one time (Wayth et al., 2018).

5.3 Results

5.3.1 Cross-power spectra

We plot the spherically averaged cross-power spectra for our six different reionization histories in Figure 5.3, showing the original and delayed-end models in the top and bottom panels, respectively. The different reionization models are shown by the coloured lines with HM12 in blue, Late in purple, and Very Late in orange. The sign of the power spectrum is indicated by the style of the line, with solid representing negative and dotted indicating positive. We note that, as denoted in Eq. 5.9, positive values of the cross-power spectrum correspond to correlation between pixels, whereas negative power corresponds to anticorrelation. The shape of the power spectrum seen in most of the panels has a widely peaked negative component at large scales (small k), and an approximately power law-shaped positive component at small scales (large k).

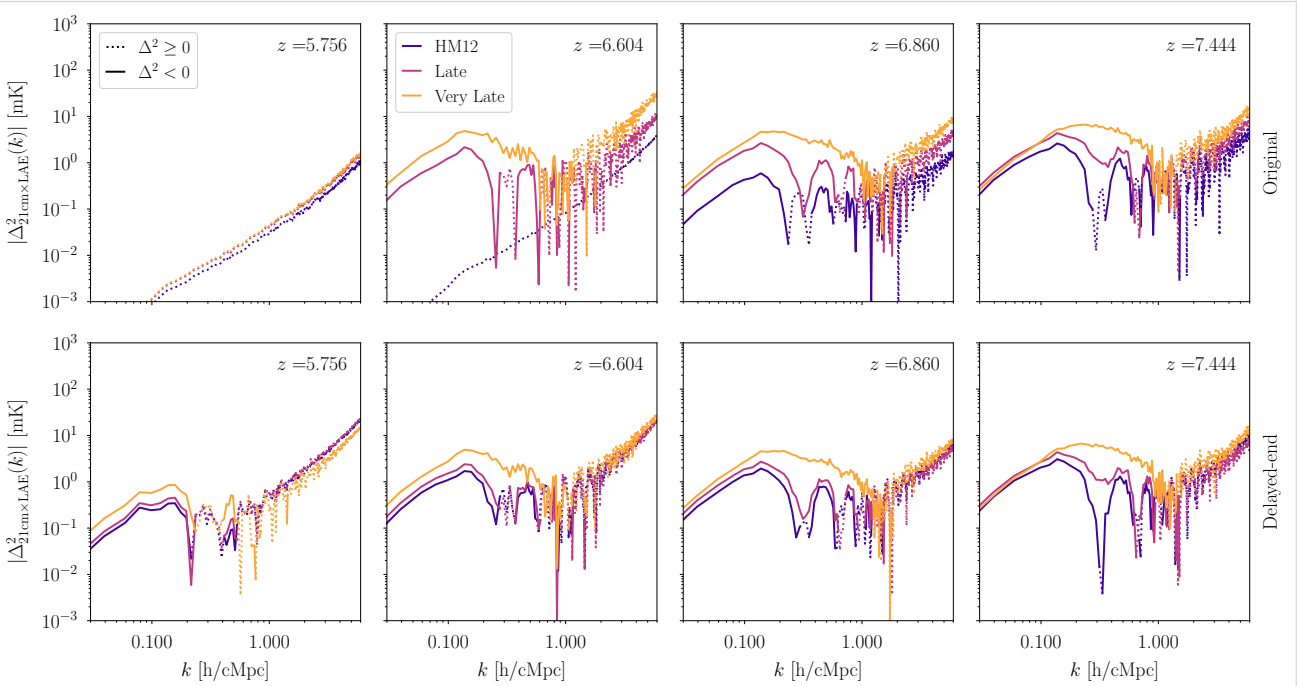


Fig. 5.3 Cross-power spectra for the reionization histories: HM12 (blue), Late (purple), Very Late (orange). The top panels show the original reionization histories, whilst the bottom panels show the delayed-end versions. From left to right the panels show increasing redshift, starting at $z = 5.756$ on the left and ending with $z = 7.444$ on the right. The dotted lines indicate where the power spectrum is positive ($\Delta^2 \geq 0$), whilst the solid lines indicate that the power spectrum is negative ($\Delta^2 < 0$).

On large scales we expect the LAEs to be anticorrelated with the 21-cm brightness, as they reside in the more massive overdensities which ionise inside-out (Iliev et al., 2006). On scales smaller than the typical bubble size we expect to see correlation, with positive power contributions from pairs of pixels in the overdensities around LAEs, where the gas self-shields and must be ionised outside-in (Miralda-Escudé et al., 2000). The general shape of our results is consistent with previous work (Furlanetto & Lidz, 2007; Lidz et al., 2009; Wiersma et al., 2013; Vrbanec et al., 2016; Kubota et al., 2018)

In more detail we see that for the original reionization histories — with reionization ending by $z = 6$ — there is no negative power in the left panel at $z = 5.756$. At these redshifts the large-scale IGM is ionised, and so we see only positive correlation resulting from the residual/self-shielded neutral gas in the overdensities that host LAEs. For the delayed-end reionization histories the $z = 5.756$ panel shows the hint of the anticorrelation signal at large scales, since there are still neutral islands of gas remaining. At the higher redshifts, for both the original and delayed-end scenarios, we see the anticorrelation signal as described above. Another feature present in many of the power spectra is an oscillatory shape around intermediate scales. This has been observed in other simulations (such as Wiersma et al., 2013; Vrbanec et al., 2016), and represents a harmonic imprinting of the typical bubble size.

Comparing the different reionization histories, we note that the Very Late model has the most power on approximately all scales, and the HM12 model has the least. This is because the amplitude of the power spectrum traces the mean neutral hydrogen fraction, with the Very Late model reionizing latest and hence having the highest mean neutral fraction at a given redshift. We see that the position of the broad peak in the anticorrelation component is dependent on the reionization history, as is the scale when the sign of the power spectrum turns over. This reflects the distribution of bubble sizes, which is a function of the average neutral fraction $\langle x_{\rm HI} \rangle$. Similarly we see that the small-scale positive correlation component is dependent on the reionization history. This dependency results from the amount of self-shielded neutral gas that is present within ionised regions, controlled by the background photoionisation rate $\Gamma_{\rm HI}$. As was similarly found in Kubota et al. (2018), we predict slightly more power at small scales than some previous work, which results from a more accurate simulation of the self-shielded gas in the ionised IGM. Our high resolution simulations are better able to resolve the small-scale overdensities, and we employ the self-shielding prescription of Chardin et al. (2018) to ensure that this effect is properly captured. However we note that our modelling assumes a uniform photoionisation rate, which likely acts to enhance the amount of self-shielded gas compared to a more realistic inhomogeneous photoionising field. At small-scales (large k), the amplitude of the cross-power spectrum is dependent on the self-shielding modelling, and hence is subject to more modelling uncertainty than the large scales. Whilst the anticorrelation peak in the power spectrum is robust to the self-shielding modelling, we note that the turnover scale depends on the interplay between the ionised bubble distribution and the amount of self-shielding within ionised regions. See the additional materials in Section 5.6.4 for more discussion of the self-shielding assumptions.

For the interested reader, the corresponding correlation functions are provided in the additional materials in Section 5.6.3.

5.3.2 Observational sensitivities

We plot our predictions for the observational sensitivities of measuring our delayed-end Very Late reionization history in Figure 5.4. The model power spectrum is shown with the black curves (solid and dotted lines indicating negative and positive power, respectively, as in Figure 5.3). The coloured lines show the error predicted for MWA (red), LOFAR (green), HERA (orange) and SKA (purple), combined with a Subaru/PFS LAE survey. The bottom panels show the predicted signal-to-noise ratios as a function of k for these surveys. We note that the sensitivity predictions are model-dependent; in Figure 5.4 we show only the predictions for our delayed-end Very Late model.

The general shape of the sensitivity curves is approximately a power law at large k, with a plateau at small k. At large k, the difference in sensitivity is determined by the different 21-cm survey array configurations. The 21-cm thermal noise on a given line-of-sight mode scales like $\propto \Omega^2/n_b(k_\perp)$, where Ω is the field of view of the interferometer, and n_b is the baseline density (see Eq. 5.33). Hence, for modes above $k \gtrsim 1$ cMpc/h^{-1}, we see that the large density of baselines of the SKA results in the

highest sensitivity. At small k, on the other hand, the survey sample variance comes into play. Here the different survey volumes limit the number of modes that can be measured. For our chosen survey parameters the limiting volume is the PFS survey rather than the 21-cm surveys (see Section 5.6.2), and hence all of the lines converge at small k. The SKA has a small field of view which results in a smaller survey volume compared to, for example, the MWA; in future cross-correlation surveys limited by the 21-cm survey volume, this will play a role at small k.

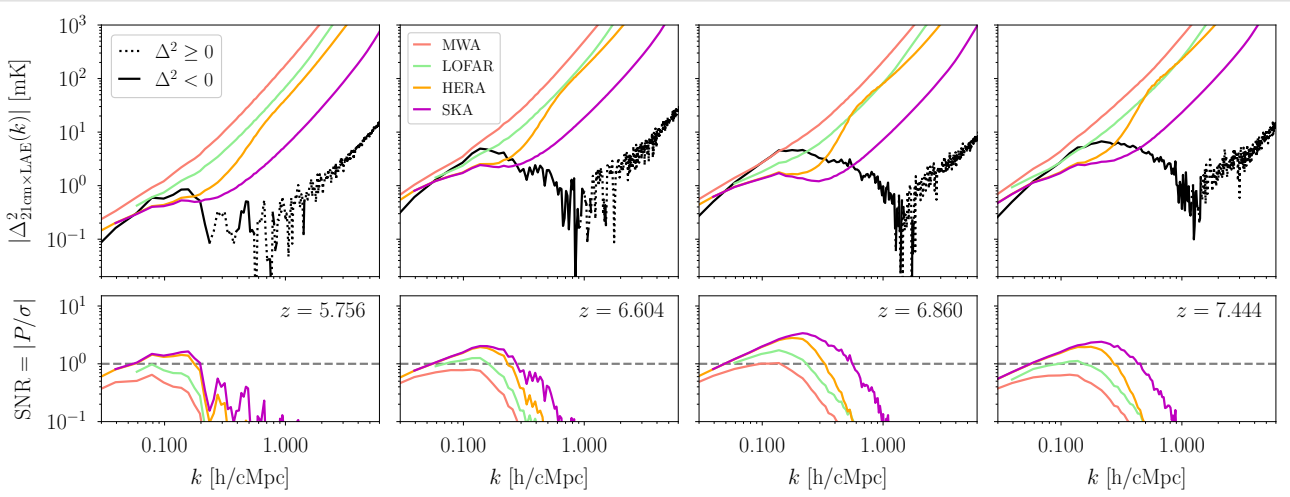


Fig. 5.4 Observational sensitivities for measuring the cross-power spectrum of the delayed-end Very Late reionization history. The coloured lines indicate different survey error predictions: red (MWA), green (LOFAR), orange (HERA) and purple (SKA). *Top panels*: The predicted 1-σ error measuring the dimensionless cross-power spectrum, with the black line indicating the underlying model power spectrum. As in Figure 5.3, from left to right the panels show increasing redshift, starting at $z = 5.756$ on the left and ending with $z = 7.444$ on the right. *Bottom panels*: the signal-to-noise ratio for each of the survey sensitivity predictions.

Our 21-cm thermal noise predictions are qualitatively consistent with those made by Kulkarni et al. (2016). Our cross-power spectrum error estimates are also broadly consistent with previous work, although we predict somewhat larger errors than those of Kubota et al. (2018). This may be due to the different parametrisation of the 21-cm surveys: for example for the SKA we assume a core radius of $r_{\mathrm{core}} = 350$ m and an outer radius of $r_{\mathrm{max}} = 6.4$ km based on the core array of the proposed V4A configuration (see Table 5.2 and Section 5.6.2), whilst Kubota et al. (2018) make estimates for a more compact configuration using $r_{\mathrm{core}} = 20$ m and $r_{\mathrm{max}} = 1$ km.

We calculate the total signal-to-noise ratios measured across linearly spaced k-bins in quadrature using,

$$\mathrm{SNR}^2_{\mathrm{total}} = \sum_i \mathrm{SNR}^2_i = \sum_i \left(\frac{P(k_i)}{\sigma(k_i)} \right)^2 , \tag{5.17}$$

where i indexes the range of k-bins between $0.1 \leq k < 10$ h/cMpc. In Table 5.3 we show the total signal-to-noise ratios for the bracketing reionization histories — the HM12 and Delayed-end Very Late models — at $z = 6.6$, using our sensitivity predictions for the four survey combinations. As expected

from Figure 5.4, the signal-to-noise ratio is highest for the PFS-SKA and PFS-HERA combinations. We also note that the Delayed-end Very Late reionization history can be detected in this k-range with a total signal-to-noise ratio greater than 5 for the PFS-SKA survey combination.

Table 5.3 Total signal-to-noise ratios for each of the survey combinations measuring the two bracketing reionization histories at $z = 6.6$, calculated in the range $0.1 \leq k < 10$ h/cMpc

Reionization	MWA	LOFAR	HERA	SKA
HM12	0.004	0.009	0.032	0.109
Delayed Very Late	1.427	2.494	4.495	5.531

In Figure 5.5 we plot a comparison of the cross-power spectrum evolution for the bracketing HM12 and Delayed-end Very Late reionization histories. We indicate the predicted 1-σ errors from a PFS-SKA observational survey using shading. In the bottom panels we show the correlation coefficient for these two models. For the redshifts near the end of reionization we see that these reionization histories could be distinguished at a level of at least 3-sigma (at large scales $k \sim 0.1$ h/cMpc).

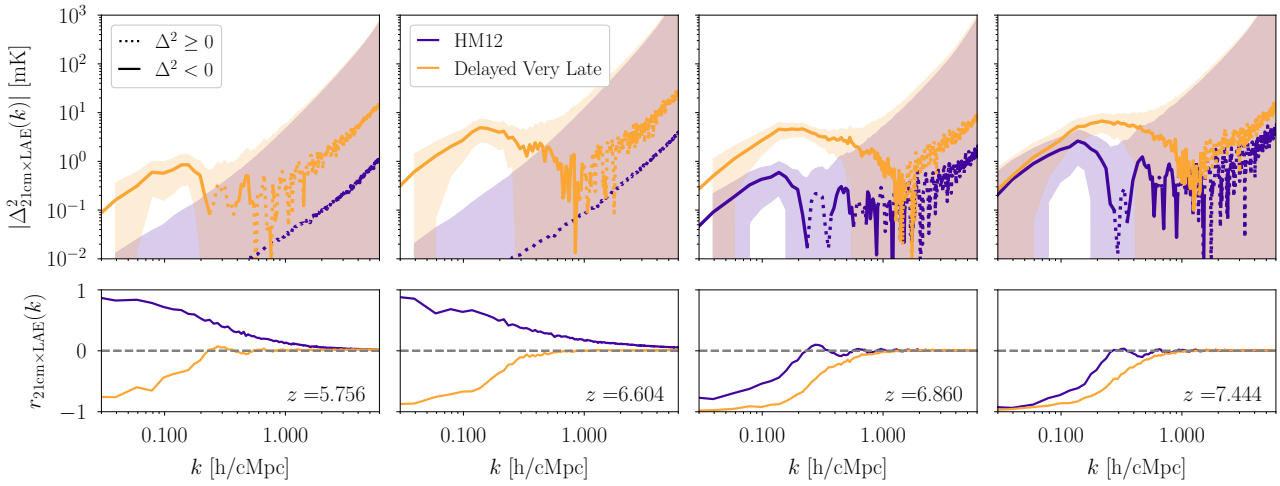


Fig. 5.5 Distinguishing reionization histories using a PFS-SKA survey: a comparison of the HM12 (purple) and delayed-end Very Late (orange) reionization histories. *Top panels:* The cross-power spectrum, with shading indicating the 1-σ error predicted for the PFS-SKA survey. *Bottom panels:* the cross-correlation coefficient. As in Figure 5.3, from left to right the panels show increasing redshift, starting at $z = 5.756$ on the left and ending with $z = 7.444$ on the right.

5.4 Discussion

5.4.1 Detecting a delayed end to reionization

The results in Figures 5.4 & 5.5 and Table 5.3 demonstrate the prospects for detecting our "Delayed-end Very Late" reionization history in the cross-correlation of 21-cm and LAE observations. The final evolution of the neutral fraction near to the end of reionization has a strong impact on the cross-correlation signal. We have compared our two bracketing reionization histories:

- **HM12**, for which reionization has finished by $z = 6.7$. This means that at the $z = 5.7$ and $z = 6.6$ narrowbands, the large-scale IGM is fully ionised and the background photoionisation is reaching spatial equilibrium.
- **Delayed-end Very Late**, for which reionization only finishes around $z = 5.3$. This means that at the two lower redshift narrowbands there is still significant neutral gas in the IGM.

We see in Figure 5.5 that it is possible to distinguish these different reionization scenarios using a PFS-SKA survey. In particular we note that for the lower redshifts the cross-power spectrum behaves very differently for these scenarios: in the HM12 case the lack of bubble structure leads to only positive power even at large scales, whilst in the delayed-end case we see the imprint of the bubble structure in the form of negative power and a clear turnover scale.

We also note in passing that our results suggest it may be possible to detect the signal from ionised bubbles in the cross-power spectrum at redshifts as low as $z = 6.6$. As has been noted previously (Dumitru et al., 2019; Lidz et al., 2009; Zahn et al., 2007; Furlanetto et al., 2004c), whilst reionization is still ongoing the typical bubble size is imprinted on the power spectrum as a turnover in the sign. In our models we find a dip in the power at this turnover scale, which means it may not be possible to separate the power spectrum signal from the noise at that scale. However, we find in all our models (at intermediate redshifts $z \sim 6.6$) that it is possible to detect the power on scales slightly larger than this turnover. For example in the Delayed-end Very Late model, using a PFS-SKA survey at $z = 6.6$, it would be possible to measure the power-spectrum at scales around $k \sim 0.2$ h/cMpc, close to the turnover at $k \sim 0.3$ h/cMpc, with a signal-to-noise ratio of ~ 3, and hence provide some constraints on the typical bubble size.

Similarly we note that the size of the negative peak in the cross-power spectrum at large scales may offer another route to constraining the progress of reionization and the typical bubble size. In Figure 5.6 we plot the magnitude of the dimensionless cross-power spectrum at this negative peak as a function of the average ionised fraction. We note that this forms a tight correlation, with a more ionised IGM resulting in a smaller peak signal. This reflects the fact that the amplitude of the cross-power spectrum traces the mean neutral fraction. As the position of the turnover scale is dependent on our self-shielding assumptions, the size of the negative peak is possibly a more robust feature for constraining the ionised bubble distribution than the turnover scale. We note however that

the sharp transition of the bubble edge in real space is more easily seen in the correlation function (see Section 5.6.3).

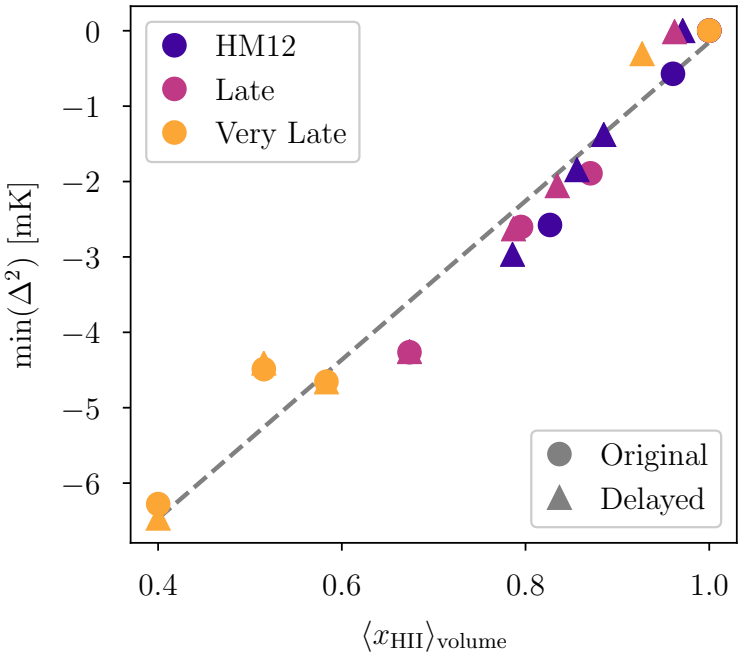


Fig. 5.6 Dependence of the cross-power spectrum features with the reionization history: the correlation between the volume-averaged ionised fraction, $\langle x_{\rm HII} \rangle$, and the size of the negative peak in the power spectrum, $\min(\Delta^2)$. The different points are derived from our six reionization histories, differentiated by the colours and symbols.

5.4.2 Future survey sensitivities

The results presented in the previous section are predictions for the established Subaru HSC Deep field, which has a field of view of ~ 27 deg^2 and a depth of $r \sim 27$ (Miyazaki et al., 2018; Ouchi et al., 2018). In Figure 5.7 we demonstrate the effect of expanding the LAE survey in either area or depth, in combination with an SKA 21-cm survey at $z = 6.6$ (left panel) and $z = 7.4$ (right panel). Increasing the total field of view of the survey reduces the sample variance by allowing more line-of-sight modes to be observed for a given k-mode. Decreasing the minimum LAE luminosity (i.e. increasing the depth of the survey) allows the survey to sample larger numbers of galaxies, and hence reduces the shot noise.

We see that a combined factor of three improvement in these quantities[1] results in approximately double the total signal-to-noise ratio. We might expect that the survey time scales approximately linearly with changes to the area (at fixed depth), but approximately quadratically with changes to the depth (at fixed area) (Djorgovski et al., 2013). Considering the regime of incremental improvements in either area or depth, we see in Figure 5.7 that a higher signal-to-noise ratio can be achieved with a

[1] That is, a threefold increase in area and a threefold decrease in the minimum luminosity.

smaller increase in observing time by increasing the survey area rather than the depth. When the Subaru/PFS survey volume is comparable to the SKA's survey volume then further improvements in area will have no effect, and hence increasing the depth is then the only option. At higher redshifts such as $z = 7.4$, we find that increasing the area of the survey (at fixed depth) provides less of an improvement compared to lower redshifts. For higher redshifts increasing both the depth and the survey area in combination is required to boost the signal-to-noise.

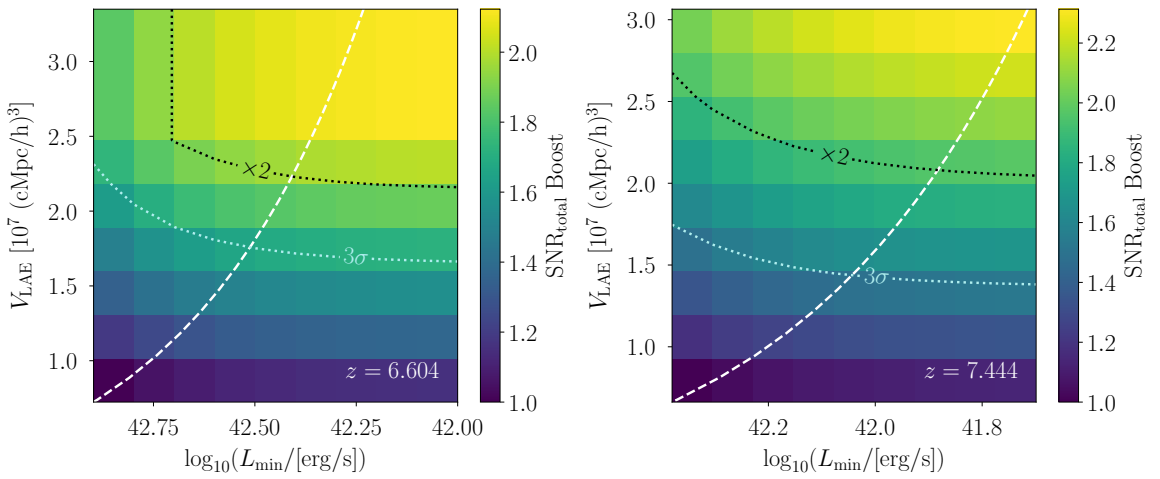


Fig. 5.7 Predictions for the boost in the total signal-to-noise ratio of a PFS-SKA survey measuring the delayed-end Very Late reionization history at $z = 6.6$ ($z = 7.4$) in the left (right) panel, calculated in the range $0.1 \leq k < 10$ h/cMpc. The PFS survey is expanded to have an increased field of view (corresponding to an increased survey volume, $V_{\rm LAE}$, shown on the vertical axis) and depth (parametrised by the minimum observed Lyα luminosity, $L_{\rm min}$, shown on the horizontal axis). The reference $\rm SNR_{total}$ is calculated for our default parametrisation of a PFS-SKA survey in which the LAE component of the survey has $A_\theta = 27$ deg^2, $V_{\rm LAE} = 2.3 \times 10^7$ cMpc^3 (2.1×10^7 cMpc^3) and $L_{\rm min} = 7.9 \times 10^{42}$ erg/s (2.4×10^{42} ergs/s). The black dotted contour indicates the values that give a boost in $\rm SNR_{total}$ of 2, whilst the white dashed line indicates proportional changes in both $V_{\rm LAE}$ and $L_{\rm min}$. The dotted light turquoise line indicates the survey parameters which give enough sensitivity to distinguish the delayed Very Late history from the HM12 history at $k = 0.2$ h/cMpc to a 3-σ level.

Plans for the WFIRST telescope include proposals for (Spergel et al., 2015):

- a High Latitude Survey (HLS) with a field of view of ~ 2000 deg^2 and a depth of 27th magnitude in Y/J/H/F184 bands,
- the option for deeper sub-fields within the HLS, with fields of view of ~ 20–50 deg^2 and depths of 28–29th magnitude.

The WFIRST's Wide Field Instrument will have both imaging and spectroscopy modes. This telescope may therefore offer further prospects for observing the 21cm-LAE cross-correlation beyond Subaru/PFS.

5.5 Conclusions

In this chapter we have employed realistic simulation models of the 21-cm signal and LAE population evolution to predict the 21cm-LAE cross-power spectrum during the epoch of reionization. Our models are calibrated to be consistent with observations of the CMB optical depth, post-reionization Lyman-α forest measurements and LAE population statistics at $z > 5$ including the luminosity function, equivalent width distribution and angular clustering function.

We confirm the predictions of recent works (Kubota et al., 2018; Hutter et al., 2018, 2017) which suggested that this cross-correlation signal can be observed with future surveys, such as a combination of Subaru/PFS and SKA. We have also found that:

- It should be possible to observe the cross-power spectrum signal for a late delayed-end reionization scenario at redshifts as low as $z = 5.7$. This scenario has been found to predict the behaviour of the post-reionization Lyman-α forest (Kulkarni et al., 2019a; Keating et al., 2020; Nasir & D'Aloisio, 2019), and is consistent with the current Planck constraints on the Thomson optical depth to the CMB (Planck Collaboration, 2018b). Targeting these lower redshifts with a Subaru/PFS-SKA cross-correlation survey will therefore allow us to constrain the manner in which reionization ended.

- It is possible to achieve better sensitivity on the cross-power spectrum by expanding the Subaru HSC survey. Increasing either the total field of view or the depth can reduce the total measurement variance, although for incremental increases we recommend increasing the survey area first. This can provide improved sensitivity until the LAE survey volume is as large as the 21-cm survey volume.

The 21cm-LAE cross-correlation signal can provide a robust constraint on the history of reionization, making possible a 21-cm-based inference that might otherwise be plagued by foregrounds. Our simulations suggest that future observations with Subaru/PFS and the SKA — at redshifts as low as $z = 5.7$ — can provide insight into this final phase transition's end.

5.6 Additional Material

5.6.1 Power spectrum estimation

We follow the approach of Smith et al. (2008) to estimate the power spectrum. Computationally we calculate the Fourier transforms of our simulation quantities using Fast Fourier Transforms (FFT), taking advantage of the `FFTW` library (Frigo & Johnson, 2005). We correct for the grid discreteness by dividing out the Fourier transform of the grid interpolation window function,

$$\tilde{\delta} = \tilde{\delta}(\mathbf{k}) / W_{\mathrm{CIC}}(\mathbf{k}), \tag{5.18}$$

where $W_{\rm CIC}(\mathbf{k})$ is the Fourier transform of the Cloud-In-Cell (CIC) window function,

$$W_{\rm CIC}(\mathbf{k}) = \left[\mathrm{sinc}\left(\frac{\pi k_1}{2k_{\rm Ny}}\right)\mathrm{sinc}\left(\frac{\pi k_2}{2k_{\rm Ny}}\right)\mathrm{sinc}\left(\frac{\pi k_3}{2k_{\rm Ny}}\right)\right]^2, \tag{5.19}$$

where $k_{\rm Ny} = \pi N_{\rm grid}/L_{\rm box}$ is the Nyquist frequency (for $N_{\rm grid}$ cells on a side of length $L_{\rm box}$).

At a scale k the spherically-averaged auto-power spectrum can be estimated as,

$$\hat{P}(k) = \frac{V}{N_m}\sum_{i=1}^{N_m}|\tilde{\delta}(\mathbf{k}_i)|^2, \tag{5.20}$$

$$= \frac{V}{N_{\rm grid}^6}\left(\frac{1}{N_m}\sum_{i=1}^{N_m}|\tilde{\delta}^{\texttt{FFTW}}(\mathbf{k}_i)|^2\right), \tag{5.21}$$

where N_m is the number of modes in a spherical shell and V is the simulation volume, $V = L_{\rm box}^3$. Note here that the FFTW transform is an unnormalised discrete Fourier transform (DFT), hence we must include the appropriate normalisation $\tilde{\delta} = \tilde{\delta}^{\texttt{FFTW}}/N_{\rm grid}^3$. Similarly for the cross-power spectrum,

$$\hat{P}_{a,b}(k) = \frac{V}{N_m}\sum_{i=1}^{N_m}\frac{[\tilde{\delta}_a^*(\mathbf{k}_i)\tilde{\delta}_b(\mathbf{k}_i) + \tilde{\delta}_a(\mathbf{k}_i)\tilde{\delta}_b^*(\mathbf{k}_i)]}{2}, \tag{5.22}$$

where $\tilde{\delta}^*$ indicates the complex conjugate.

We correct for shot noise when calculating power spectrum, $P_{\rm LAE}(k)$, following Smith et al. (2008),

$$\hat{P}(k) = \hat{P}(k) - \hat{P}^{\rm shot}(k). \tag{5.23}$$

For the LAE auto-power spectrum the shot noise term is given by,

$$\hat{P}^{\rm shot}_{\rm LAE\times LAE}(k) = \frac{V}{N_{\rm obj}}, \tag{5.24}$$

where $N_{\rm obj}$ is the number of objects (e.g. $N_{\rm LAE}$) in the simulation volume. Note the 21cm-LAE cross-power spectrum estimator does not need to be corrected for LAE shot noise. The shot noise corrections affect small scales, $k > 1$ h/cMpc, and are strongly dependent on the number of observed LAEs, $N_{\rm LAE}$.

5.6.2 Survey sensitivities

In this section we derive expressions for the noise terms used in Eqs. (5.11) & (5.12).

21cm surveys

We follow the treatments of Parsons et al. (2012), Geil et al. (2011) and McQuinn et al. (2006) in order to derive the thermal noise of 21-cm observations. We neglect any systematic errors introduced by foreground removal (and refer readers to Yoshiura et al., 2018, for further exploration of foreground effects). The variance in a measurement of the 21-cm power spectrum is given by,

$$\mathrm{var}[P_{21\mathrm{cm}}(k,\mu)] = \left[P_{21\mathrm{cm}}(k,\mu) + P_{21\mathrm{cm}}^{\mathrm{noise}}(k,\mu)\right]^2, \tag{5.25}$$

where the first term on the right-hand side is the contribution from sample variance (i.e. random error), and the second term is the contribution from thermal noise in the interferometer (i.e. systematic error).

Interferometers measure fringe visibilities, $V(\mathbf{U}, \nu)$, where $\mathbf{U} = (u, v)$ is the image coordinate measured in wavelengths (sometimes called the *spatial frequency*), and ν is the frequency of the measurement. Note we are assuming the flat-sky approximation in which we neglect line-of-sight dependence. We will use the notation $\tilde{V}(u, v, \eta)$ to refer to the frequency Fourier transform of the visibility (η is an inverse frequency). The observed power spectrum of brightness temperature fluctuations can be approximately related to $\tilde{V}$ by (see Parsons et al., 2012, for a detailed derivation),

$$P(k,\mu) = \langle|\tilde{T}_b(\mathbf{k})|^2\rangle \simeq \left[\left(\frac{\lambda^2}{2k_B}\right)^2 \frac{D^2\,\Delta D}{\Omega\,B^2}\right] \tilde{V}^2(u, v, \eta), \tag{5.26}$$

where λ is the observed wavelength, k_B is Boltzmann's constant, D is the comoving distance to the surveyed emission, ΔD is the comoving radial depth of the survey, B is the bandwidth of the measurement, and Ω is the field of view of the telescope given by $\Omega = \lambda^2/A_e$, where A_e is the effective area of an antenna. The Fourier mode (u, v, η) measured by the inteferometer is related to the cosmological $\mathbf{k}$-mode by,

$$(Dk_x, Dk_y, \Delta Dk_z/B) = 2\pi(u, v, \eta), \tag{5.27}$$

or equivalently,

$$Dk_\perp = 2\pi U, \quad \Delta Dk_\parallel/B = 2\pi\eta, \tag{5.28}$$

where we have split the $\mathbf{k}$-mode into components parallel and perpendicular to the line-of-sight of the observation, $\mathbf{k} = \mathbf{k}_\perp + \mathbf{k}_\parallel$, and $U = |\mathbf{U}|$.

We wish to derive the thermal noise of a given k-mode that is left *after* attempting to subtract the noise power (using any available information). In time $t_\mathbf{k}$ we can make $2Bt_\mathbf{k}$ independent measurements of the *observation-independent* system temperature, T_{sys}. Assuming Gaussian random fluctuations in our measurements of the system temperature (which have magnitude $\sqrt{2}T_{\mathrm{sys}}$), the thermal fluctuations in the detector have *observation-dependent* rms brightness temperature,

$$(T_{\mathrm{rms}}^{\mathrm{noise}})^2 = \frac{(\sqrt{2}T_{\mathrm{sys}})^2}{2Bt_\mathbf{k}} = \frac{T_{\mathrm{sys}}^2}{Bt_\mathbf{k}}, \tag{5.29}$$

which is also known as the *radiometer equation*. This adds a white-noise contribution to the rms amplitude of the Fourier transform of the visibility given by,

$$\tilde{V}^{\rm noise}(u, v, \eta) = \frac{2k_B}{\lambda^2}\Omega\, B\, T^{\rm noise}_{\rm rms}(u, v, \eta). \tag{5.30}$$

Using Eq. (5.26) the noise power is therefore given by (Parsons et al., 2012; McQuinn et al., 2006; Morales, 2005),

$$P^{\rm noise}_{21\rm cm}(k, \mu) = D^2 \Delta D \frac{\Omega\, T^2_{\rm sys}}{2\, B\, t_{\mathbf{k}}}, \tag{5.31}$$

where we have introduced a factor of two in the denominator to account for using two orthogonal polarisation measurements to measure the total signal.

We note that $t_{\mathbf{k}}$ is the average time that a given $\mathbf{k}$-mode can be observed by the interferometer (not simply the total integration time), which will depend explicitly on the configuration and position (latitude) of a given array. Here we estimate this time as (McQuinn et al., 2006),

$$t_{\mathbf{k}} \simeq t_{\rm int}\, n_b(k_\perp)/\Omega, \tag{5.32}$$

where $t_{\rm int}$ is the total integration time of the observation and $n_b(k_\perp)$ is the number density of baselines observing at a given time that sample a particular transverse k-mode, $k_\perp = k\sqrt{1-\mu^2}$, which accounts for the antenna array geometry.

The final expression for the thermal noise on a (k, μ)-mode is therefore given by,

$$P^{\rm noise}_{21\rm cm}(k, \mu) = \frac{D^2 \Delta D}{n_b(k_\perp)} \frac{\Omega^2\, T^2_{\rm sys}}{2\, B\, t_{\rm int}}. \tag{5.33}$$

We assume a system temperature dominated by the sky temperature (Wyithe & Morales, 2007),

$$T_{\rm sys} \simeq 280{\rm K} \left(\frac{1+z}{7.5}\right)^{2.3}. \tag{5.34}$$

We note that the comoving survey depth is approximately given by (Parsons et al., 2012),

$$\Delta D \simeq 1.7 \left(\frac{B}{0.1\ {\rm MHz}}\right) \left(\frac{1+z}{10}\right)^{0.5} \left(\frac{\Omega_{\rm m} h^2}{0.15}\right)^{-1.5} {\rm cMpc}. \tag{5.35}$$

For all telescopes we assume an integration time of $t_{\rm int} = 1000$ hrs, a bandpass of $B = 8$ MHz (which corresponds to depth of $\Delta D(z = 7) \simeq 100$ cMpc), and a frequency resolution of $\Delta\nu = 50$ kHz. And finally we note that using the above quantities the 21-cm survey volume is given by,

$$V_{\rm survey} = D^2 \Delta D\, \Omega = D^2 \Delta D \left(\frac{\lambda^2}{A_e}\right). \tag{5.36}$$

We list the 21-cm survey parameters that we use to estimate sensitivities in Table 5.2.

The baseline number density, $n_b(k_\perp)$, is equivalently parametrised by U (following Eq. (5.28)), such that we can also write it as $n_b(U)$. This density is normalised so that the integral of $n_b(U)\mathrm{d}U$ over the observed half-plane is the total number of baselines,

$$N_{\text{baselines}} = \frac{1}{2} N_a (N_a - 1), \tag{5.37}$$

where N_a is the total number of antennae[2] in the array. If we approximate that the array is circularly symmetric, we can calculate the number density of baselines as a convolution (Geil et al., 2011; Datta et al., 2007),

$$n_b(U) = C_b \int_0^{r_{\max}} \mathrm{d}r\, 2\pi r\, n_a(r) \int_0^{2\pi} \mathrm{d}\phi\, n_a(|\mathbf{r} - \lambda \mathbf{U}|), \tag{5.38}$$

where n_a is the number density of antennae, and C_b is a (frequency-dependent) normalisation constant that ensures the integral over the half-plane is given by Eq. (5.37),

$$N_{\text{baselines}} = \int_0^{U_{\max}} \mathrm{d}U\, 2\pi\, U\, n_b(U). \tag{5.39}$$

Using Eq. (5.28) we can rewrite,

$$|\mathbf{r} - \lambda \mathbf{U}| = \sqrt{r^2 - 2\chi r k_\perp \cos\phi + \chi^2 k_\perp^2}, \tag{5.40}$$

where $\chi = \lambda U / k_\perp = \lambda D / 2\pi$.

We can perform the convolution in Eq. (5.38) numerically using the actual array configuration, $n_a(r)$, of a given interferometer, or alternatively using an idealised model. In this work we use an idealised model following the configuration described in Geil et al. (2011),

$$n_a(r) = \begin{cases} n_{\text{core}} & r \le r_{\text{core}}, \\ n_{\text{core}} \left(r_{\text{core}}/r\right)^2 & r_{\text{core}} < r < r_{\max}, \\ 0 & \text{otherwise}, \end{cases} \tag{5.41}$$

where n_{core} is the number density of a core region within radius r_{core}, given by,

$$n_{\text{core}} = \frac{N_a}{\pi r_{\text{core}}^2 [1 + 2 \ln (r_{\max}/r_{\text{core}})]}. \tag{5.42}$$

The number density in the core region is limited by the physical size of the antenna tiles, such that $n_{\text{core}} < 1/A_e$. We choose values for the parameters N_a, r_{core} and $r_{\max}$ that reflect the configurations of MWA-Phase II (extended configuration), LOFAR, HERA and SKA1-low, similar to those of Kulkarni et al. (2016). These values can be found in Table 5.2.

[2] In practice each "antenna tile" is a collection of individual antenna elements, such as a phased-array of dipole antennae. This collection is often referred to simply as an antenna, or a station.

LAE surveys

We follow the treatments of Feldman et al. (1994) and Tegmark et al. (1997) in order to derive errors on the LAE galaxy survey. As for the 21-cm case, the variance in a measurement of the LAE power spectrum is given by,

$$\mathrm{var}[P_{\mathrm{LAE}}(k,\mu)] = \left[P_{\mathrm{LAE}}(k,\mu) + P_{\mathrm{LAE}}^{\mathrm{noise}}(k,\mu)\right]^2, \tag{5.43}$$

where as before the first term on the right-hand side is the contribution from sample variance, and the second term is the contribution from galaxy shot noise.

If we include the Poissonian shot noise term explicitly, Eq. (5.43) can be rewritten (Feldman et al., 1994) (dropping the k, μ labels),

$$\frac{\mathrm{var}[P_{\mathrm{LAE}}]}{P_{\mathrm{LAE}}} = \left[1 + \frac{1}{\bar{n}_{\mathrm{LAE}}\, P_{\mathrm{LAE}}}\right]^2, \tag{5.44}$$

where $\bar{n}_{\mathrm{LAE}}$ is the mean LAE number density in the observed sample.

We assume the redshift estimates (photometric from HSC or spectroscopic from PFS) introduce a Gaussian spread in the redshift-space measurement of n_{LAE} along the line-of-sight (Loureiro et al., 2019),

$$n_{\mathrm{LAE}}(z) = \int \mathrm{d}z'\, n_{\mathrm{LAE}}^{\mathrm{int}}(z - z') \exp\left(-\frac{z'^2}{2\sigma_z^2}\right), \tag{5.45}$$

where $n_{\mathrm{LAE}}^{\mathrm{int}}(z)$ is the underlying, intrinsic redshift distribution of the LAEs, and σ_z is the typical redshift measurement error. This distorts the measured power spectrum (estimated in redshift space),

$$P(k,\mu) \longrightarrow P(k,\mu)\, \exp\left(-\left[k_\parallel c\sigma_z / H(z)\right]^2\right), \tag{5.46}$$

where $k_\parallel = \mu k$ is the component of the wave-vector parallel to the line-of-sight, c is the speed of light, and $H(z)$ is the Hubble parameter. Including this effect in Eq. (5.44), we find that the shot noise term is increased by a factor $\exp\left(\left[k_\parallel c\sigma_z / H(z)\right]^2\right)$.

The final shot noise error term for a (k,μ)-mode of the LAE galaxy survey is therefore given by,

$$P_{\mathrm{LAE}}^{\mathrm{noise}}(k,\mu) = \frac{\exp\left(\left[k_\parallel c\sigma_z / H(z)\right]^2\right)}{\bar{n}_{\mathrm{LAE}}}. \tag{5.47}$$

As outlined in section 5.2.5, we estimate the sensitivity of a survey similar to SILVERRUSH (Ouchi et al., 2018), performed using the Subaru telescope. We can estimate $\bar{n}_{\mathrm{LAE}}$ for a given depth by integrating the luminosity functions observed in the SILVERRUSH survey; in practice we use the Schechter fits from Itoh et al. (2018). We choose $\sigma_z = 0.0007$ for the PFS redshift errors (Takada et al., 2014), following Kubota et al. (2018).

Volume matching and available modes

In practice a cross-correlation study between 21-cm and LAE surveys would require careful volume matching to ensure that the same physical region of the Universe is being sampled by each telescope. In order to replicate this in our sensitivity estimates we limit the included modes of the surveys to that of the more restrictive survey, and also set $V_{\rm survey}$ in Eq. (5.13) to be the smaller of the two volumes $\{V_{21\rm cm}, V_{\rm LAE}\}$. For the telescopes considered in this work, and for a given $\mathbf{k}$-mode, we note that the μ-modes available for extraction are limited by the 21-cm surveys (Furlanetto & Lidz, 2007). In particular there is a maximum parallel component set by the depth of the survey,

$$\mu_{\rm max} = \min(1, k/k_{\parallel,\rm min}), \tag{5.48}$$

and a minimum transverse component given by,

$$\mu^2_{\rm min} = \max(0, 1 - k^2_{\perp,\rm max}/k^2), \tag{5.49}$$

where $k_{\parallel,\rm min} = 2\pi/\Delta D$ and $k_{\perp,\rm max} = 2\pi u_{\rm max}/D$, for maximum baseline spatial frequency $u_{\rm max} = |\mathbf{u}|_{\rm max}$.

In Table 5.2 we list the survey volumes and available modes for each of the surveys considered in this work at $z = 6.6$. The LAE survey volumes are calculated assuming a wide-field narrowband geometry,

$$V_{\rm LAE} \simeq d(z)^3 A_\theta \left(\frac{d(z+\Delta z)}{d(z)} - 1 \right), \tag{5.50}$$

where $d(z)$ is the comoving distance to a redshift z, the narrowband depth is assumed to be $\Delta z = 0.1$, and the angular area of the survey is $A_\theta = 27$ deg^2 (Subaru Deep field). The 21cm volumes are calculated using Eq. (5.36).

5.6.3 Cross-correlation functions

We can use the k-space cross-power spectra to derive the real-space correlation functions by taking the Fourier transform,

$$\xi(r) = \frac{1}{(2\pi)^3} \int P(k) \mathrm{sinc}(kr) 4\pi k^2 \mathrm{d}k, \tag{5.51}$$

where $\mathrm{sinc}(x) = \sin(x)/x$. We note that the integrand in this transform is highly oscillatory, and so the numerical integration must be performed with care. In Figure 5.8 we plot the correlation function evolution for all our reionization histories (as in Figure 5.3 for the cross-power spectra). These results are in qualitative agreement with previous work, such as Hutter et al. (2018, 2017); Kubota et al. (2018)[3].

[3] Although we note that when making comparisons, one must be careful of units. In particular we correlate the 21-cm brightness temperature itself (and not the contrast from the mean), and hence retain units of mK in the cross-correlation function.

The correlation functions can be thought of as the average 21-cm brightness profiles around LAEs. At the intermediate redshifts we find that there is an anticorrelation signal for scales $r < 10$ cMpc/h, which dies off at larger scales. This is driven by the presence of ionised bubbles around LAEs, which sit in the overdensities that power reionization. We note that it might be possible to use the scale at which the correlation function dies to zero to estimate the typical ionised bubble size; further exploration of this is left to future work

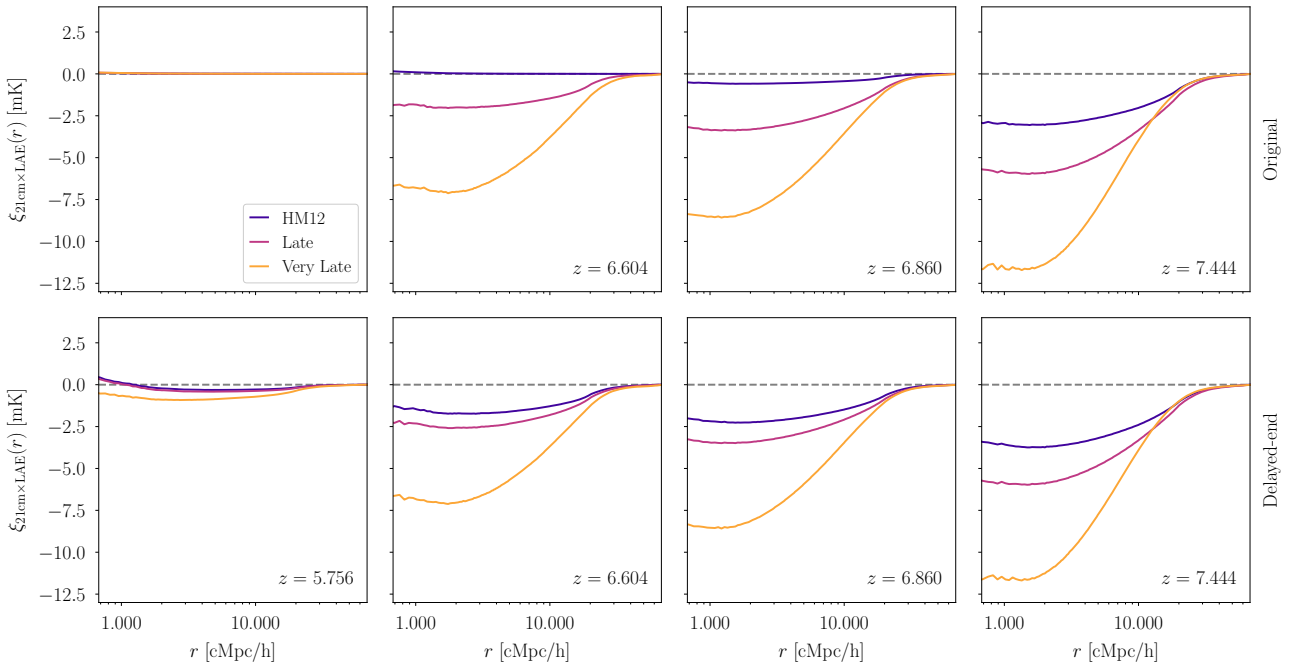


Fig. 5.8 Cross-correlation functions for the reionization histories: HM12 (blue), Late (purple), Very Late (orange). The top panels show the original reionization histories, whilst the bottom panels show the delayed-end versions. From left to right the panels show increasing redshift, starting at $z = 5.756$ on the left and ending with $z = 7.444$ on the right. The dashed horizontal line indicates $\xi(r) = 0$.

Conversely at very small scales, we see the hint of an upturn in the correlation signal due to residual/self-shielded neutral hydrogen in the overdensities that host the LAEs. In the $z = 5.756$ panels we see that the original models show no correlation signal, whilst the delayed-end models show a slight anti-correlation at intermediate scales. Reionization has not fully completed in these models, and hence the ionised bubbles have not fully percolated.

In Figure 5.9 we plot a comparison of the correlation functions for the HM12 and delayed-end Very Late models, showing the predicted 1-σ errors for a PFS-SKA survey as in Figure 5.5. The errors have been propagated numerically using a Monte Carlo sampling of the power spectrum, assuming Gaussian statistics. We note that the chosen observational limits (see Table 5.1) result in a larger number of LAEs at $z = 6.9$ than at the lower redshifts, which reduces the shot noise at that redshift, and hence the 1-σ errors are narrower in the $z = 6.9$ panel.

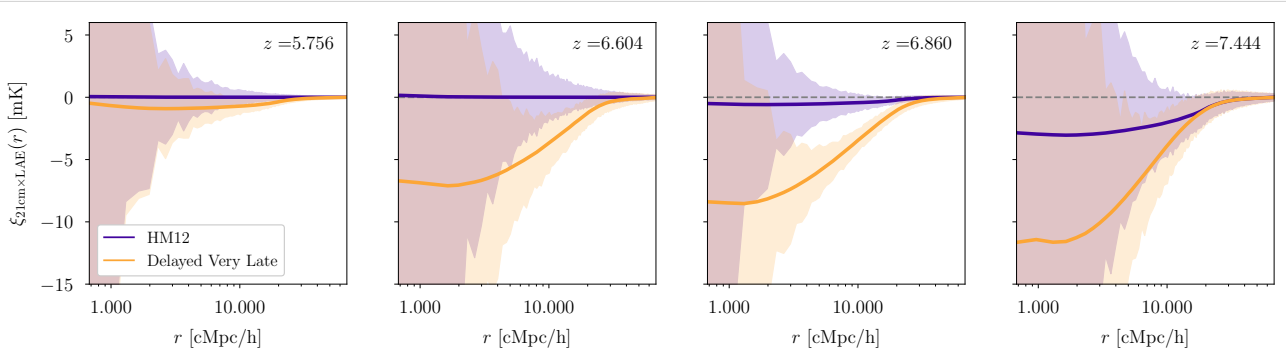


Fig. 5.9 As in Figure 5.5, distinguishing reionization histories using a PFS-SKA survey: a comparison of the HM12 (purple) and delayed-end Very Late (orange) reionization histories. Here we plot the cross-correlation function, with shading indicating the 1-σ error predicted for the PFS-SKA survey.

5.6.4 The effect of self-shielding

In this section we explore the dependence of our results on the assumed self-shielding prescription. The small-scale correlations between LAEs and the 21-cm brightness depends on the amount of neutral gas in the vicinities of host haloes. The extent of the self-shielding in gas close to the LAE can effect both how bright the gas is in 21-cm emission, but also how much the Lyα emission from the galaxies is attenuated. Hence if there is more self-shielding we can expect to see an increase in the correlation signal[4].

In Figure 5.10 we test the effect of varying our self-shielding prescription on the cross-power spectrum and cross-correlation function. We note that the self-shielding prescription affects both the 21-cm brightness calculation and the LAE transmission, so there is a coupled change in both of our observables. In this plot we show our fiducial self-shielding prescription derived from Chardin et al. (2018) in red, the Rahmati et al. (2013) prescription in green, and using no prescription in blue. Considering the power spectrum we see that on large scales (below $k < 0.3$ h/cMpc), varying the self-shielding assumptions does not have any effect. As we found in section 5.3, current and upcoming surveys are most sensitive to the reionization signal on these large scales, hence our forecasts for measuring the 21-cm LAE cross-power spectrum are robust to the self-shielding modelling choices. However on smaller scales (above $k \geq 0.3$ h/cMpc) we see that the self-shielding modelling has a large effect on the resulting power spectrum. In particular, although across all prescriptions the shape of the power spectrum remains a power law with similar index, the amplitude is a strong function of the self-shielding. As a result, the turnover scale is also dependent on the self-shielding modelling. In the right panel of Figure 5.10 we show the cross-correlation function. Here we see that the amount of self-shielding controls the extent to which the correlation function turns over at small scales.

[4]As we found in Chapters 3 & 4, the brightest LAEs are found in the most massive haloes, where we also find significant amounts of self-shielded neutral gas. If we increase the amount of self-shielding then only the brightest LAEs – whose positions are highly correlated with the bright self-shielded gas – remain visible, thus enhancing the clustering.

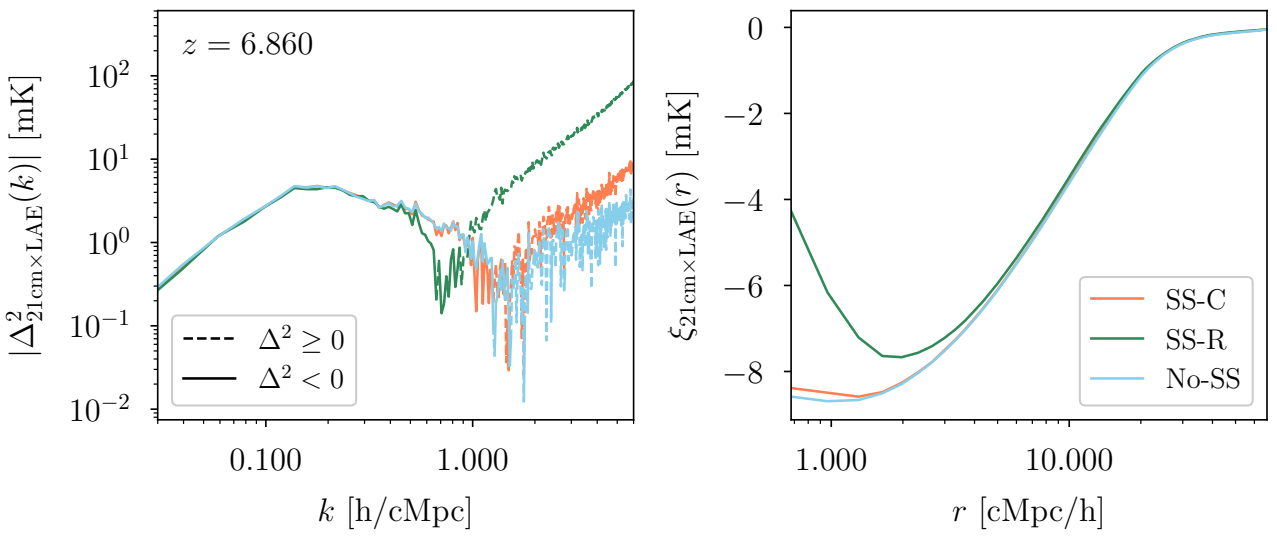


Fig. 5.10 The effect of self-shielding on the cross-power spectrum (left, as in Figure 5.3) and cross-correlation function (right, as in Figure 5.8) at $z = 6.860$ for the Very Late reionization history. We show our fiducial self-shielding prescription derived from Chardin et al. (2018) in red, the Rahmati et al. (2013) prescription in green, and using no self-shielding prescription in blue.

Chapter 6

Conclusions and Outlook

6.1 Summary

In this thesis we have made significant progress investigating the Epoch of Reionization (EoR) using numerical simulations combined with modelling of Lyman-α emitting galaxies (LAEs) and the 21-cm signal. We have successfully navigated a number of unique challenges which arise from both a theoretical and observational point of view when trying to understand the EoR.

The numerical experiments performed in this work employed the cutting edge Sherwood simulation suite to capture the cosmological hydrodynamic evolution of the intergalactic medium (IGM). We modelled the transitioning ionisation state of the IGM by post-processing these simulations with a semi-numerical photon counting scheme (also referred to as an "excursion-set" method), with careful calibration of a background photoionisation rate to ensure self-consistency with the emissivity of the ionising sources. Using these simulations of a reionizing IGM, in Chapter 3 we explored the host-halo dependence of the LAE attenuation. In Chapter 4 we took our modelling a step further to populate the simulations with a mock population of LAEs, allowing us to investigate the effect of reionization on LAE population statistics such as the luminosity function. Finally in Chapter 5 we introduced a model for the 21-cm signal, and combined this with our LAE population modelling in order to make forecasts for the sensitivity of upcoming surveys to measuring the LAE-21cm cross-correlation signal.

In each of these chapters we made comparisons with recent observational results in order to infer which of our reionization scenarious are plausible. Our simulations successfully reproduce a variety of observations, and thus provide insight into the evolution of the high-redshift LAE population. The main conclusions of this work can be summarised as follows:

- We have updated the understanding of how observations of LAEs and the 21-cm signal are sensitive to the reionization of the IGM, and how they can therefore provide excellent probes of this distant epoch, as suggested by other authors (such as Haiman, 2002; Furlanetto et al., 2004a,b; McQuinn et al., 2006; Dijkstra et al., 2007; Dayal et al., 2009; Laursen et al., 2009; Bolton

& Haehnelt, 2013; Jensen et al., 2013; Hutter et al., 2014; Jensen et al., 2014; Mesinger et al., 2015; Choudhury et al., 2015; Kubota et al., 2018). In Chapter 3 we found that the transmission of Lyman-α emission from high redshift galaxies is significantly attenuated by neutral gas in the IGM, and that different reionization histories will result in different evolution of this transmission. This conclusion was extended in the work in Chapter 4, where our results have advanced the understanding of how the evolution of LAE population statistics is transformed by the progress of reionization. Finally in Chapter 5 we expanded our modelling to include the 21-cm signal, which can be cross-correlated with LAE observations to overcome difficulties with foreground emission, and confirmed that this cross-correlation is also sensitive to reionization.

- We have firmly established that a relatively "late" end to reionization is favoured by our LAE predictions. In Chapter 3 we tested three bracketing reionization histories which varied the end-point, $z_{\rm end}$, and gradient, $\mathrm{d}\langle x_{\rm HII}(z)\rangle/\mathrm{d}z$, of the reionization epoch. Our results for the relative transmission evolution of LAEs suggest that "early" reionization scenarios such as the model of Haardt & Madau (2012) are inconsistent with current observations. In Chapter 4 we tested the impact of reionization histories which end as late as $z \sim 5.3$, including a scenario which has the same neutral fraction evolution, $\langle x_{\rm HI}(z)\rangle$, as the model of Kulkarni et al. (2019a) that successfully predicts the opacity fluctuations in the Lyman-α forest and is consistent with the Planck Collaboration (2018b) measurement of the Thomson optical depth of the Cosmic Microwave Background (CMB). We found that the evolution of the LAE luminosity function and equivalent width distribution predicted by this late reionization history is consistent with current observations from Subaru surveys (Konno et al., 2018; Itoh et al., 2018; Ota et al., 2017; Zheng et al., 2017; Konno et al., 2014). As in Chapter 3 we found that the earlier Haardt & Madau (2012) history is inconsistent with observations, being unable to capture the rapid decline in the space density of LAEs beyond $z > 6.6$. We note that these conclusions agree with recent independent LAE analyses such as Mason et al. (2018a); Inoue et al. (2018).

  As previously mentioned, recent observations of the statistics of the Lyman-α forest (Becker et al., 2018, 2015b) suggest a large scatter in opacity at high redshifts around $z \gtrsim 5$ which are well explained by large residual neutral "islands" that would be leftover in late reionization scenarios (Kulkarni et al., 2019a; Keating et al., 2020; Nasir & D'Aloisio, 2019). Over the past decade the measured value of the electron scattering optical depth to the CMB has come down from $\tau \sim 0.17$ (measured with WMAP, by Hinshaw et al., 2013) to $\tau \sim 0.05$ (measured with Planck, by Planck Collaboration, 2018b), which also supports the conclusion that reionization may have ended relatively abruptly around $z \sim 5$.

- When modelling LAEs and the 21-cm signal in Chapters 3, 4 & 5 we found that the LAE and 21-cm reionization observables are not only sensitive to the distribution of ionised regions, but that they are also influenced by the presence of recombined and residual neutral gas *within* ionised regions. This means that theoretical models must be careful to ensure they capture

self-shielding and the partial ionisation within ionised bubbles; conversely it also complicates the inference of the average progress of reionization from observations. We found in Chapter 3 that the relative importance of gas within ionised regions is a function of the local environment, for example the region of infalling gas around an LAE host halo. These conclusions build on the independent work of Sadoun et al. (2017). We note that in some cases, such as in the LAE-21cm cross-power spectrum, this complication can be evaded by restricting consideration to large scale modes which are more sensitive to the ionised bubble distribution.

Alongside these main outcomes we can also draw some further incidental conclusions, summarised as follows:

- In Chapter 3 we found that differing gas kinematics around low and high mass haloes can impact Lyman-α transmission, and that this might therefore contribute to a differing visibility of faint and bright LAEs. We tested this further on our mock LAE populations in Chapter 4 and found that on average the Lyman-α transmission fraction does indeed increase with LAE brightness. We note that observational studies such as Zheng et al. (2017) have reported a bump in the bright-end of the LAE luminosity function, whilst Stark et al. (2017) found that the Lyman-α fraction of LBG galaxies evolved differently for bright and faint samples.

- In Chapter 4 we found that the clustering of LAEs is significantly enhanced by reionization. In particular, for the physically motivated choice of LBG duty cycle, our mock LAE population overpredicts the angular correlation function at higher redshifts compared to the current SILVERRUSH results (Ouchi et al., 2018). The observed angular correlation shows little evolution between $z = 5.7$ and $z = 6.6$, which seems to be at odds with the evolution in the luminosity function and equivalent width distribution.

- In Chapter 5 we found that the signal-to-noise ratio that surveys can achieve when measuring the LAE-21cm cross-power spectrum can be boosted by expanding the galaxy survey area and depth. We forecasted sensivities for a collection of current and upcoming surveys; our fidicual predictions used a parametrisation based on the (current) SILVERRUSH galaxy survey. We found this survey configuration to be the limiting factor, and that in particular expanding the survey area would reduce the total measurement variance on large scale modes.

6.2 Future Research Outlook

Many aspects of the EoR and high redshift galaxy evolution remain uncertain. Fortunately the future looks bright for this research field, with a number of promising observational surveys planned, as well as the continued advancement of theoretical and numerical modelling in part made possible by the progress of computer technology.

On the observational side, the launch of upcoming telescopes such as the James Webb Space Telescope (JWST, Park et al., 2020) and Wide Field Infrared Survey Telescope[1] (WFIRST, Koekemoer et al., 2019) will greatly expand our ability to detect high-redshift galaxies. The Square Kilometer Array (SKA) radio interferometer will also allow us to detect the 21-cm signal deep into reionization (Koopmans et al., 2015). Conversely on the theoretical side, the 2020s are set to be the era of exascale computing, with new heterogeneous architectures allowing computers to push to greater speeds (Goz et al., 2020). Alongside this there is continued algorithmic advancement taking advantage of new hardware to run bigger and better simulations (Cavelan et al., 2020).

Reionization

Although our understanding of reionization has evolved considerably over the past decades, it remains to confirm:

- The exact contributions of galaxies, quasars and other sources to the ionising photon budget required to ionise the IGM. We must still determine how the ionising escape fraction of galaxies evolves, and whether therefore galaxies are indeed the dominant component of the ionising budget.
- The exact reionization history — *when did reionization start, how long did it take, and when did it finish?* As we summarised in Section 6.1, there is mounting evidence suggesting an end to reionization as late as $z \sim 5.3$, but the early stages are relatively unknown.

To finally answer these questions with certainty, the future promises advances in both our observational and theoretical capabilities.

Proposed surveys using the Near Infrared Camera (NIRCam) and Near Infrared Spectrograph (NIRSpec) on the JWST will study the evolution of galaxies to redshifts as high as $z \sim 15$ (Rieke et al., 2019). Extremely Large Telescopes (ELTs) will also offer the next stage of photometrically identified galaxies, providing follow-up of deep WFIRST imaging surveys to allow us to better map out reionization (Finkelstein, 2019). We note there are also promising synergies between future galaxy surveys and 21-cm observations (Zackrisson et al., 2020). Improved measurements of the CMB may also provide further constraints on reionization. For example Roy et al. (2020) predict that future surveys such as CMB-S4 (Abazajian & et al.,, 2019) and PICO (Hanany et al., 2019) combined with 21-cm observations with the HERA radio interferometer (DeBoer et al., 2017) would be capable of detecting the cross-correlation between the Thomson optical depth, τ_e, and the 21-cm signal, providing information about the evolution of reionization.

Recent, cutting-edge simulations of reionization include SPHINX (Rosdahl et al., 2018) and CoDA II (Ocvirk et al., 2020), both of which employ radiative transfer coupled to hydrodynamics on-the-fly at high enough resolutions to capture galaxy evolution and the ionisation of the IGM by starlight.

[1] We note that WFIRST was renamed the Nancy Grace Roman Space Telescope (NGRST) on 20 May 2020.

Simulations such as these can be used to gain greater insight into the impacts of an inhomogeneous reionization on the key observables such as the CMB, Lyman-α forest, LAEs and the 21-cm signal (for example, Gronke et al., 2020). Furthermore, they can hopefully provide a more detailed picture of the physical processes controlling the escape of ionising radiation from galaxies (Katz et al., 2019b).

LAEs

The LAE model described in this thesis offers a very effective theoretical tool for progessing our interpretations of LAE observations. We note that alongside this thesis work, the model has been employed in the work of Keating et al. (2020)[2] which investigated correlations between "holes" in LAE surface brightness maps and the large troughs found in the Lyman-α forest (Becker et al., 2018). Furthermore, work is currently being undertaking using the modelling from this thesis to explore: (i) the cross-correlation between LAEs and Lyman-α intensity maps of the IGM at high redshifts (Witstok et al., 2019), (ii) the cross-correlation between galaxies (both LAEs and LBGs) and the thermal Sunyaev-Zeldovich (tSZ) distortion of the CMB. Similarly to the cross-correlation with the 21-cm signal described in Chapter 5, the use of LAE observations in combination with other observables can be a powerful probe of reionization and high-redshift phenomena.

In general, research into LAEs will benefit from the upcoming observational surveys and numerical simulations described above. In particular it will be important in future work to revisit some of the assumptions made in this thesis, such as the simple model for the intrinsic LAE emission profile employed in Chapters 3, 4 and 5, as well as the role of the CGM in the escape of radiation from high-redshift galaxies. Improved modelling of the emission profiles of LAEs can be achieved by post-processing simulations with Lyman-α radiative transfer codes (Smith et al., 2018). Highly parallelised codes such as Rascas (Michel-Dansac et al., 2020) can be used to efficiently post-process high-resolution simulations (Kimm et al., 2019) in order to analyse the emission profiles of representative samples of simulated galaxies. These improved models will be matched by the latest observations with telescopes such as the Very Large Telescope (VLT), whose integral field spectrograph – the multi-unit spectroscopic explorer (MUSE) – can achieve the high spatial and spectral resolution to properly capture the emission of high-redshift galaxies (Smit et al., 2018).

Alongside our improving understanding of the statistics of high-redshift LAEs as a population, there has been considerable recent activity trying to understand and use observations of individual ultraluminous LAEs found deep within the reionization epoch. Three interesting examples are COLA1 ($z \sim 6.6$, Matthee et al., 2018), NEPLA4 ($z \sim 6.6$, Songaila et al., 2018) and MACS1149-JD1 ($z \sim 9.1$, Hashimoto et al., 2018), all of which have been observed with emission bluewards of Lyman-α — suggesting they must reside in large ionised regions. Mason & Gronke (2020) constructed an analytic model to try to infer the size of cosmological H II regions from the galaxy properties and this presence of blue emission. Another promising window into high-redshift LAEs is provided by the proximity zone of quasars; Bosman et al. (2019) identified three LAEs with double-peaked emission profiles

[2] Keating et al. (2020) was undertaken in collaboration with this author.

close to quasar J0836 which can provide insight into the ionising properties of both galaxies and quasars.

21-cm Signal

Surveys with current interferometer arrays such as LOFAR (Mertens et al., 2020) and MWA (Li et al., 2019; Barry et al., 2019) continue to improve their measurements of the 21-cm auto-power spectrum, with upper bounds already making it possible to rule out parts of the (astrophysical) modelling parameter space (Ghara et al., 2020). In particular these current upper bounds put limits on how far the spin temperature can deviate from the saturated limit (discussed briefly in Section 1.2.2) and thus rule out some of the more exotic scenarios for the physics of cosmic dawn and reionization. Alongside this, studies continue to try to understand and interpret the EDGES measurement (for example Tauscher et al., 2020; Leo et al., 2020). As discussed in Chapter 5, the upcoming SKA telescope will greatly expand our ability to detect the high-redshift 21-cm signal (Greig et al., 2020).

Parallel to the advances in our understanding of reionization, the 21-cm signal promises to reveal more about the cosmology of the early universe. For example the global 21-cm may be an effective probe of the primordial power spectrum (Yoshiura et al., 2020). Finally we note that beyond the global 21-cm signal and the power spectrum, the bispectrum is also a potential source of reionization information (Hutter et al., 2020).

⋆ ⋆ ⋆

"*You know the greatest danger facing us is ourselves, an irrational fear of the unknown. But there's no such thing as the unknown –* ***only things temporarily hidden, temporarily not understood.***"

— James T. Kirk

References

Aarseth S. J., 1963, MNRAS, 126, 223

Aarseth S. J., 1966, MNRAS, 132, 35

Aarseth S. J., 1971, Ap&SS, 14, 20

Abazajian K., et al., 2019, arXiv e-prints, p. arXiv:1907.04473

Abbott T. M. C., et al., 2018, MNRAS, 480, 3879

Abel T., Bryan G. L., Norman M. L., 2002, Science, 295, 93

Ahn C. P., et al., 2012, ApJS, 203, 21

Alam S., et al., 2017, MNRAS, 470, 2617

Alpher R. A., Bethe H., Gamow G., 1948, Physical Review, 73, 803

Alvarez M. A., et al., 2019, arXiv e-prints, p. arXiv:1903.04580

Ando M., Ohta K., Iwata I., Akiyama M., Aoki K., Tamura N., 2007, PASJ, 59, 717

Appel A. W., 1985, SIAM Journal on Scientific and Statistical Computing, 6, 85

Aubert D., Teyssier R., 2008, MNRAS, 387, 295

Aubert D., Teyssier R., 2010, ApJ, 724, 244

Avery L. W., House L. L., 1968, ApJ, 152, 493

Bañados E., et al., 2018, Nature, 553, 473

Bagley M. B., et al., 2017, ApJ, 837, 11

Bardeen J. M., Steinhardt P. J., Turner M. S., 1983, Phys. Rev. D, 28, 679

Barkana R., Loeb A., 2001, Phys. Rep., 349, 125

Barnes J., Hut P., 1986, Nature, 324, 446

Barnes L. A., Haehnelt M. G., Tescari E., Viel M., 2011, MNRAS, 416, 1723

Barry N., et al., 2019, ApJ, 884, 1

Bauer A., Springel V., 2012, MNRAS, 423, 2558

Beane A., Lidz A., 2018, ApJ, 867, 26

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

Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096

Becker G. D., Bolton J. S., Lidz A., 2015a, PASA, 32, e045

Becker G. D., Bolton J. S., Madau P., Pettini M., Ryan-Weber E. V., Venemans B. P., 2015b, MNRAS, 447, 3402

Becker G. D., Davies F. B., Furlanetto S. R., Malkan M. A., Boera E., Douglass C., 2018, ApJ, 863, 92

Bekenstein J. D., 1974, Phys. Rev. D, 9, 3292

Bengaly C. A. P., Maartens R., Randriamiarinarivo N. r., Baloyi A., 2019, J. Cosmology Astropart. Phys., 2019, 025

Bertschinger E., 1998, ARA&A, 36, 599

Bhowmick A. K., Campbell D., DiMatteo T., Feng Y., 2018, arXiv e-prints, p. arXiv:1806.10612

Bludman S. A., 1998, ApJ, 508, 535

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

Bolton J. S., Haehnelt M. G., 2007, MNRAS, 374, 493

Bolton J. S., Haehnelt M. G., 2013, MNRAS, 429, 1695

Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897

Boltzmann L., 1872, K. Acad. Wiss.(Wein) Sitzb., II Abt, 66

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

Bosman S. E. I., Kakiichi K., Meyer R. A., Gronke M., Laporte N., Ellis R. S., 2019, arXiv e-prints, p. arXiv:1912.11486

Bouwens R. J., Illingworth G. D., Franx M., Ford H., 2008, ApJ, 686, 230

Bouwens R. J., et al., 2015, ApJ, 803, 34

Bowman J. D., Rogers A. E. E., Monsalve R. A., Mozdzen T. J., Mahesh N., 2018, Nature, 555, 67

Boyle L. A., Steinhardt P. J., Turok N., 2006, Phys. Rev. Lett., 96, 111301

Bromm V., Larson R. B., 2004, ARA&A, 42, 79

Caldwell R. R., Dave R., Steinhardt P. J., 1998, Phys. Rev. Lett., 80, 1582

Calverley A. P., Becker G. D., Haehnelt M. G., Bolton J. S., 2011, MNRAS, 412, 2543

Cavelan A., Cabezón R. M., Grabarczyk M., Ciorba F. M., 2020, arXiv e-prints, p. arXiv:2005.02656

Chabanier S., Millea M., Palanque-Delabrouille N., 2019, MNRAS, 489, 2247

Chandrasekhar S., 1960, Radiative transfer. Dover Publications

Chang T. C., Gong Y., Santos M., Silva M. B., Aguirre J., Doré O., Pritchard J., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 4 (arXiv:1501.04654)

Chapman E., Santos M. G., 2019, MNRAS, p. 2288

Chapman E., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 5 (arXiv:1501.04429)

Chardin J., Haehnelt M. G., Aubert D., Puchwein E., 2015, MNRAS, 453, 2943

Chardin J., Puchwein E., Haehnelt M. G., 2017, MNRAS, 465, 3429

Chardin J., Kulkarni G., Haehnelt M. G., 2018, MNRAS,

Charlot S., Fall S. M., 1993, ApJ, 415, 580

Chen N., Doussot A., Trac H., Cen R., 2020, arXiv e-prints, p. arXiv:2004.07854

Choudhury T. R., 2009, Current Science, 97, 841

Choudhury T. R., Paranjape A., 2018, MNRAS, 481, 3821

Choudhury T. R., Ferrara A., Gallerani S., 2008, MNRAS, 385, L58

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

Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, MNRAS, 452, 261

Ciardi B., Ferrara A., Governato F., Jenkins A., 2000, Monthly Notices of the Royal Astronomical Society, 314, 611

Clarke C., Carswell B., 2014, Principles of Astrophysical Fluid Dynamics. Cambridge University Press

Colombi S., Dodelson S., Widrow L. M., 1996, ApJ, 458, 1

Cooley J. W., Lewis P. A. W., Peter Welch D., 1967, IEEE Trans. Audio Electroacoust

Couchman H. M. P., 1991, ApJ, 368, L23

Couchman H. M. P., Rees M. J., 1986, MNRAS, 221, 53

Coupon J., et al., 2012, A&A, 542, A5

Coupon J., Czakon N., Bosch J., Komiyama Y., Medezinski E., Miyazaki S., Oguri M., 2018, PASJ, 70, S7

Curtis-Lake E., et al., 2012, MNRAS, 422, 1425

Datta K. K., Bharadwaj S., Choudhury T. R., 2007, MNRAS, 382, 809

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Dawoodbhoy T., et al., 2018, MNRAS, 480, 1740

Dayal P., Ferrara A., 2012, MNRAS, 421, 2568

Dayal P., Ferrara A., 2018, Phys. Rep., 780, 1

Dayal P., Ferrara A., Saro A., Salvaterra R., Borgani S., Tornatore L., 2009, MNRAS, 400, 2000

DeBoer D. R., et al., 2017, PASP, 129, 045001

Dehnen W., Read J. I., 2011, European Physical Journal Plus, 126, 55

Diener C., et al., 2017, MNRAS, 471, 3186

Dijkstra M., 2014, PASA, 31, e040

Dijkstra M., Westra E., 2010, MNRAS, 401, 2343

Dijkstra M., Wyithe J. S. B., 2012, MNRAS, 419, 3181

Dijkstra M., Haiman Z., Loeb A., 2004, ApJ, 613, 646

Dijkstra M., Haiman Z., Spaans M., 2006, ApJ, 649, 14

Dijkstra M., Lidz A., Wyithe J. S. B., 2007, MNRAS, 377, 1175

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

Djorgovski S. G., Mahabal A., Drake A., Graham M., Donalek C., 2013, Sky Surveys. p. 223, doi:10.1007/978-94-007-5618-2_5

Dopita M. A., Krauss L. M., Sutherland R. S., Kobayashi C., Lineweaver C. H., 2011, Ap&SS, 335, 345

Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium

Dumitru S., Kulkarni G., Lagache G., Haehnelt M. G., 2019, MNRAS, 485, 3486

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

East W. E., Wojtak R., Abel T., 2018, Phys. Rev. D, 97, 043509

Eckhardt R., 1987, Los Alamos Sci., Special Issue, 15, 131

Eddington A. S., 1916, MNRAS, 77, 16

Efstathiou G., 1986, Nature, 322, 311

Efstathiou G., 1988, Effects of reionization on microwave background anisotropies.. pp 299–319

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

Efstathiou G., Eastwood J. W., 1981, MNRAS, 194, 503

Efstathiou G., Gratton S., 2020, arXiv e-prints, p. arXiv:2002.06892

Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241

Egan C. A., Lineweaver C. H., 2010, ApJ, 710, 1825

Einstein A., 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin, pp 844–847

Eisenstein D. J., Hu W., Silk J., Szalay A. S., 1998, ApJ, 494, L1

Erb D. K., Pettini M., Shapley A. E., Steidel C. C., Law D. R., Reddy N. A., 2010, ApJ, 719, 1168

Erb D. K., et al., 2014, ApJ, 795, 33

Erkal D., 2013, PhD thesis, The University of Chicago

Fan X., et al., 2006, AJ, 132, 117

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

Feldman H. A., Kaiser N., Peacock J. A., 1994, ApJ, 426, 23

Feng C., Cooray A., Keating B., 2017, preprint, (arXiv:1701.07005)

Ferland G. J., 1983, Hazy, A Brief Introduction to Cloudy 33

Ferland G. J., et al., 2017, Rev. Mexicana Astron. Astrofis., 53, 385

Fialkov A., 2018, in Jelić V., van der Hulst T., eds, IAU Symposium Vol. 333, Peering towards Cosmic Dawn. pp 22–25, doi:10.1017/S1743921318000194

Field G. B., 1959, ApJ, 129, 536

Finkelstein S. L., 2019, in American Astronomical Society Meeting Abstracts #234. p. 315.03

Finkelstein S. L., et al., 2015, ApJ, 810, 71

Finkelstein S. L., et al., 2019, ApJ, 879, 36

Finlator K., Oppenheimer B. D., Davé R., 2011, MNRAS, 410, 1703

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

Finlator K., et al., 2017, MNRAS, 464, 1633

Friedmann A., 1922, Zeitschrift fur Physik, 10, 377

Frigo M., Johnson S. G., 2005, Proceedings of the IEEE, 93, 216

Fumagalli M., O'Meara J. M., Prochaska J. X., 2016, MNRAS, 455, 4100

Furlanetto S. R., Lidz A., 2007, ApJ, 660, 1030

Furlanetto S. R., Sokasian A., Hernquist L., 2004a, MNRAS, 347, 187

Furlanetto S. R., Hernquist L., Zaldarriaga M., 2004b, MNRAS, 354, 695

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004c, ApJ, 613, 1

Furlanetto S. R., Zaldarriaga M., Hernquist L., 2004d, ApJ, 613, 16

Furlanetto S. R., Schaye J., Springel V., Hernquist L., 2005, ApJ, 622, 7

Furlanetto S. R., Oh S. P., Briggs F. H., 2006, Phys. Rep., 433, 181

Gaikwad P., et al., 2020, arXiv e-prints, p. arXiv:2001.10018

Gawiser E., et al., 2007, ApJ, 671, 278

Geil P. M., Gaensler B. M., Wyithe J. S. B., 2011, MNRAS, 418, 516

Ghara R., Choudhury T. R., Datta K. K., 2015, MNRAS, 447, 1806

Ghara R., et al., 2020, MNRAS, 493, 4728

Giallongo E., et al., 2015, A&A, 578, A83

Gingold R. A., Monaghan J. J., 1982, Journal of Computational Physics, 46, 429

Giri S. K., Mellema G., Dixon K. L., Iliev I. T., 2018a, MNRAS, 473, 2949

Giri S. K., Mellema G., Ghara R., 2018b, MNRAS, 479, 5596

Gnedin N. Y., 1995, ApJS, 97, 231

Gnedin N. Y., 2000a, ApJ, 535, 530

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

Gnedin N. Y., Abel T., 2001, New A, 6, 437

González M., Audit E., Huynh P., 2007, A&A, 464, 429

Goz D., Tornatore L., Taffoni G., Murante G., 2020, in Ballester P., Ibsen J., Solar M., Shortridge K., eds, Astronomical Society of the Pacific Conference Series Vol. 522, Astronomical Data Analysis Software and Systems XXVII. p. 45

Greengard L., 1988, PhD thesis, -

Greig B., Mesinger A., Haiman Z., Simcoe R. A., 2017, MNRAS, 466, 4239

Greig B., Mesinger A., Koopmans L. V. E., 2020, MNRAS, 491, 1398

Greiner J., et al., 2009, ApJ, 693, 1610

Gronke M., Dijkstra M., 2016, ApJ, 826, 14

Gronke M., Bull P., Dijkstra M., 2015, ApJ, 812, 123

Gronke M., et al., 2020, arXiv e-prints, p. arXiv:2004.14496

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

Gurung-López S., Orsi Á. A., Bonoli S., Baugh C. M., Lacey C. G., 2019, MNRAS, 486, 1882

Guth A. H., 1981, Phys. Rev. D, 23, 347

Haardt F., Madau P., 2012, ApJ, 746, 125

Haehnelt M. G., Steinmetz M., 1998, MNRAS, 298, L21

Haehnelt M. G., Madau P., Kudritzki R., Haardt F., 2001, ApJ, 549, L151

Haiman Z., 2002, ApJ, 576, L1

Haiman Z., Knox L., 1999, Reionization of the Intergalactic Medium and its Effect on the CMB. p. 227

Hanany S., et al., 2019, in BAAS. p. 194 (arXiv:1908.07495)

Harrison E. R., 1970, Phys. Rev. D, 1, 2726

Hashimoto T., et al., 2015, ApJ, 812, 157

Hashimoto T., et al., 2018, Nature, 557, 392

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

Hawking S. W., 1975, Communications in Mathematical Physics, 43, 199

Hawking S. W., 1982, Physics Letters B, 115, 295

Hawking S. W., Ellis G. F. R., 1973, The large-scale structure of space-time.

Heitmann K., et al., 2019, ApJS, 245, 16

Hernquist L., Katz N., Weinberg D. H., Miralda-Escudé J., 1996, ApJ, 457, L51

Herring J. R., 1979, in Durst F., Launder B. E., Schmidt F. W., Whitelaw J. H., eds, Turbulent Shear Flows I. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 347–352

Hills R., Kulkarni G., Meerburg P. D., Puchwein E., 2018, Nature, 564, E32

Hinshaw G., et al., 2013, ApJS, 208, 19

Hockney R. W., Eastwood J. W., 1988, Computer simulation using particles

Holmberg E., 1941, ApJ, 94, 385

Hoyle F., 1994, Home is where the wind blows : chapters from a cosmologist's life / Fred Hoyle.. University Science Books, Mill Valley, Calif.

Hu E. M., Cowie L. L., McMahon R. G., 1998, ApJ, 502, L99

Hu E. M., Cowie L. L., Barger A. J., Capak P., Kakazu Y., Trouille L., 2010, ApJ, 725, 394

Hu E. M., Cowie L. L., Songaila A., Barger A. J., Rosenwasser B., Wold I. G. B., 2016, ApJ, 825, L7

Hubble E., 1929, Proceedings of the National Academy of Science, 15, 168

Hui L., Gnedin N. Y., 1997, MNRAS, 292, 27

Hutter A., Dayal P., Partl A. M., Müller V., 2014, MNRAS, 441, 2861

Hutter A., Dayal P., Müller V., 2015, MNRAS, 450, 4025

Hutter A., Dayal P., Müller V., Trott C. M., 2017, ApJ, 836, 176

Hutter A., Trott C. M., Dayal P., 2018, MNRAS, 479, L129

Hutter A., Watkinson C. A., Seiler J., Dayal P., Sinha M., Croton D. J., 2020, MNRAS, 492, 653

Iliev I. T., Mellema G., Pen U. L., Merz H., Shapiro P. R., Alvarez M. A., 2006, Monthly Notices of the Royal Astronomical Society, 369, 1625

Inoue A. K., et al., 2018, PASJ,

Itoh R., et al., 2018, ApJ, 867, 46

Izotov Y. I., Thuan T. X., 1998, ApJ, 500, 188

James R. A., Sellwood J. A., 1978, MNRAS, 182, 331

Jeans J. H., 1902, Philosophical Transactions of the Royal Society of London Series A, 199, 1

Jensen H., Laursen P., Mellema G., Iliev I. T., Sommer-Larsen J., Shapiro P. R., 2013, MNRAS, 428, 1366

Jensen H., Hayes M., Iliev I. T., Laursen P., Mellema G., Zackrisson E., 2014, MNRAS, 444, 2114

Kaiser N., 1984, ApJ, 284, L9

Kaiser N., 1987, MNRAS, 227, 1

Kakiichi K., Dijkstra M., Ciardi B., Graziani L., 2016, MNRAS, 463, 4019

Kashikawa N., et al., 2006, ApJ, 648, 7

Kashikawa N., et al., 2011, ApJ, 734, 119

Katz H., Ricotti M., 2014, MNRAS, 444, 2377

Katz H., et al., 2019a, arXiv e-prints, p. arXiv:1905.11414

Katz H., Kimm T., Haehnelt M. G., Sijacki D., Rosdahl J., Blaizot J., 2019b, Monthly Notices of the Royal Astronomical Society, 483, 1029

Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736

Kennicutt Robert C. J., 1998, ApJ, 498, 541

Khostovan A. A., et al., 2019, MNRAS, 489, 555

Kilby J. S., 1976, IEEE Transactions on Electron Devices, 23, 648

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

Klypin A. A., Shandarin S. F., 1983, MNRAS, 204, 891

Koekemoer A., et al., 2019, BAAS, 51, 550

Kogut A., et al., 2003, ApJS, 148, 161

Konno A., et al., 2014, ApJ, 797, 16

Konno A., et al., 2018, PASJ, 70, S16

Koopmans L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 1 (arXiv:1505.07568)

Kubota K., Yoshiura S., Takahashi K., Hasegawa K., Yajima H., Ouchi M., Pindor B., Webster R. L., 2018, MNRAS, 479, 2754

Kubota K., Inoue A. K., Hasegawa K., Takahashi K., 2019, arXiv e-prints, p. arXiv:1910.02361

Kuhlen M., Faucher-Giguère C.-A., 2012, MNRAS, 423, 862

Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2016, MNRAS, 463, 2583

Kulkarni G., Choudhury T. R., Puchwein E., Haehnelt M. G., 2017, MNRAS, 469, 4283

Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019a, MNRAS, 485, L24

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

Landy S. D., Szalay A. S., 1993, ApJ, 412, 64

Laursen P., Sommer-Larsen J., Andersen A. C., 2009, ApJ, 704, 1640

Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, ApJ, 728, 52

Laursen P., Sommer-Larsen J., Milvang-Jensen B., Fynbo J. P. U., Razoumov A. O., 2019, A&A, 627, A84

Lee K.-S., Giavalisco M., Conroy C., Wechsler R. H., Ferguson H. C., Somerville R. S., Dickinson M. E., Urry C. M., 2009, ApJ, 695, 368

Lemaître G., 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Leo M., Theuns T., Baugh C. M., Li B., Pascoli S., 2020, J. Cosmology Astropart. Phys., 2020, 004

Levermore C. D., 1984, J. Quant. Spec. Radiat. Transf., 31, 149

Lewis A., 2008, Phys. Rev. D, 78, 023002

Li W., et al., 2019, ApJ, 887, 141

Lidz A., Zahn O., Furlanetto S. R., McQuinn M., Hernquist L., Zaldarriaga M., 2009, ApJ, 690, 252

Lifshitz E. M., Khalatnikov I. M., 1964, Soviet Physics Uspekhi, 6, 495

Linde A. D., 1982, Physics Letters B, 108, 389

Liu A., Tegmark M., 2011, Phys. Rev. D, 83, 103006

Liu A., Parsons A. R., Trott C. M., 2014, Phys. Rev. D, 90, 023019

Loeb A., Barkana R., 2001, ARA&A, 39, 19

Loureiro A., et al., 2019, Monthly Notices of the Royal Astronomical Society, 485, 326

Lyman T., 1906, ApJ, 23, 181

Macpherson H. J., Price D. J., Lasky P. D., 2019, Phys. Rev. D, 99, 063522

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

Madau P., Rees M. J., 2000, ApJ, 542, L69

Madau P., Meiksin A., Rees M. J., 1997, ApJ, 475, 429

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

Madau P., Rees M. J., Volonteri M., Haardt F., Oh S. P., 2004, ApJ, 604, 484

Mainali R., et al., 2018, MNRAS, p. 1569

Majumdar S., Mellema G., Datta K. K., Jensen H., Choudhury T. R., Bharadwaj S., Friedrich M. M., 2014, MNRAS, 443, 2843

Mason C. A., Gronke M., 2020, arXiv e-prints, p. arXiv:2004.13065

Mason C. A., Trenti M., Treu T., 2015, ApJ, 813, 21

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2018a, ApJ, 856, 2

Mason C. A., et al., 2018b, ApJ, 857, L11

Matthee J., Sobral D., Santos S., Röttgering H., Darvish B., Mobasher B., 2015, MNRAS, 451, 400

Matthee J., Sobral D., Darvish B., Santos S., Mobasher B., Paulino- Afonso A., Röttgering H., Alegre L., 2017, MNRAS, 472, 772

Matthee J., Sobral D., Gronke M., Paulino-Afonso A., Stefanon M., Röttgering H., 2018, A&A, 619, A136

McDonald P., et al., 2006, ApJS, 163, 80

McGreer I. D., Mesinger A., D'Odorico V., 2015, MNRAS, 447, 499

McKee C. F., Tan J. C., 2008, ApJ, 681, 771

McQuinn M., Upton Sanderbeck P. R., 2016, MNRAS, 456, 47

McQuinn M., Zahn O., Zaldarriaga M., Hernquist L., Furlanetto S. R., 2006, ApJ, 653, 815

McQuinn M., Lidz A., Zahn O., Dutta S., Hernquist L., Zaldarriaga M., 2007a, MNRAS, 377, 1043

McQuinn M., Hernquist L., Zaldarriaga M., Dutta S., 2007b, MNRAS, 381, 75

Meiksin A. A., 2009, Reviews of Modern Physics, 81, 1405

Meiksin A., Madau P., 1993, ApJ, 412, 34

Mertens F. G., et al., 2020, MNRAS, 493, 1662

Mesinger A., Furlanetto S., 2007, Astrophys. J., 669, 663

Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, MNRAS, 446, 566

Metropolis N., 1985, Monte-Carlo: In the Beginning and Some Great Expectations. p. 62, doi:10.1007/BFb0049035

Michel-Dansac L., Blaizot J., Garel T., Verhamme A., Kimm T., Trebitsch M., 2020, A&A, 635, A154

Mirabel I. F., Dijkstra M., Laurent P., Loeb A., Pritchard J. R., 2011, A&A, 528, A149

Miralda-Escudé J., 2003, ApJ, 597, 66

Miralda-Escudé J., Rees M. J., 1994, Monthly Notices of the Royal Astronomical Society, 266, 343

Miralda-Escudé J., Haehnelt M., Rees M. J., 2000, ApJ, 530, 1

Mitra S., Choudhury T. R., Ferrara A., 2018, MNRAS, 473, 1416

Miyazaki S., et al., 2002, PASJ, 54, 833

Miyazaki S., et al., 2012, in Ground-based and Airborne Instrumentation for Astronomy IV. p. 84460Z, doi:10.1117/12.926844

Miyazaki S., et al., 2018, PASJ, 70, S1

Mo H., van den Bosch F. C., White S., 2010, Galaxy Formation and Evolution

Monaghan J. J., Gingold R. A., 1983, Journal of Computational Physics, 52, 374

Monaghan J. J., Lattanzio J. C., 1985, A&A, 149, 135

Moore G. E., 1998, Proceedings of the IEEE, 86, 82

Morales M. F., 2005, ApJ, 619, 678

Morales M. F., Hewitt J., 2004, ApJ, 615, 7

Moriwaki K., Yoshida N., Eide M. B., Ciardi B., 2019, arXiv e-prints, p. arXiv:1906.10863

Moukalled F., Mangani L., Darwish M., 2016, The Finite Volume Method in Computational Fluid Dynamics, 1 edn. 113, Springer International Publishing, The address

Murray S. G., Power C., Robotham A. S. G., 2013, Astronomy and Computing, 3, 23

Nakajima K., Fletcher T., Ellis R. S., Robertson B. E., Iwata I., 2018, MNRAS, 477, 2098

Nasir F., D'Aloisio A., 2019, arXiv e-prints, p. arXiv:1910.03570

Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72

Neben A. R., Stalder B., Hewitt J. N., Tonry J. L., 2017, ApJ, 849, 50

Neufeld D. A., 1990, ApJ, 350, 216

Ocvirk P., et al., 2020, MNRAS,

Ono Y., et al., 2012, ApJ, 744, 83

Orsi A., Lacey C. G., Baugh C. M., 2012, MNRAS, 425, 87

Ostriker J. P., Vishniac E. T., 1986, ApJ, 306, L51

Ota K., et al., 2017, ApJ, 844, 85

Ouchi M., et al., 2008, ApJS, 176, 301

Ouchi M., et al., 2010, ApJ, 723, 869

Ouchi M., et al., 2018, PASJ, 70, S13

Pallottini A., Ferrara A., Gallerani S., Vallini L., Maiolino R., Salvadori S., 2017, MNRAS, 465, 2540

Panamarev T., Just A., Spurzem R., Berczik P., Wang L., Arca Sedda M., 2019, MNRAS, 484, 3279

Park J., Kim H.-S., Wyithe J. S. B., Lacey C. G., 2014, MNRAS, 438, 2474

Park J., Gillet N., Mesinger A., Greig B., 2020, MNRAS, 491, 3891

Parsons A., Pober J., McQuinn M., Jacobs D., Aguirre J., 2012, ApJ, 753, 81

Partridge R. B., Peebles P. J. E., 1967, ApJ, 147, 868

Pawlik A. H., Schaye J., van Scherpenzeel E., 2010, Keeping the Universe Ionised: Photoheating and the High-redshift Clumping Factor of the Intergalactic Gas. p. 230

Peebles P. J., 1966, Phys. Rev. Lett., 16, 410

Peebles P. J. E., 1968, ApJ, 153, 1

Peebles P. J. E., 1980, The large-scale structure of the universe

Peebles P. J. E., 1982, ApJ, 263, L1

Pentericci L., et al., 2014, ApJ, 793, 113

Perlmutter S., et al., 1999, ApJ, 517, 565

Planck Collaboration 2014, A&A, 571, A27

Planck Collaboration 2016, A&A, 596, A108

Planck Collaboration 2018a, arXiv e-prints, p. arXiv:1807.06205

Planck Collaboration 2018b, preprint, p. arXiv:1807.06209 (arXiv:1807.06209)

Planck Collaboration et al., 2014, A&A, 571, A16

Pontzen A., et al., 2008, MNRAS, 390, 1349

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

Pritchard J. R., Loeb A., 2012, Reports on Progress in Physics, 75, 086901

Puchwein E., Haardt F., Haehnelt M. G., Madau P., 2019, MNRAS, 485, 47

Quinn T., Katz N., Stadel J., Lake G., 1997, arXiv e-prints, pp astro–ph/9710043

Rahmati A., Pawlik A. H., Raičevič M., Schaye J., 2013, MNRAS, 430, 2427

Rees M. J., 1968, ApJ, 153, L1

Reichardt C. L., 2016, Observing the Epoch of Reionization with the Cosmic Microwave Background. p. 227, doi:10.1007/978-3-319-21957-8_8

Reid B. A., et al., 2010, MNRAS, 404, 60

Rieke M., et al., 2019, BAAS, 51, 45

Riess A. G., et al., 1998, AJ, 116, 1009

Riess A. G., Casertano S., Yuan W., Macri L. M., Scolnic D., 2019, ApJ, 876, 85

Rindler W., 1956, MNRAS, 116, 662

Robertson H. P., 1932, Science, 76, 221

Robertson B. E., et al., 2013, ApJ, 768, 71

Robitaille T. P., 2011, A&A, 536, A79

Rosdahl J., et al., 2018, MNRAS, 479, 994

Roy A., Lapi A., Spergel D., Basak S., Baccigalupi C., 2020, J. Cosmology Astropart. Phys., 2020, 062

Rybicki G. B., Lightman A. P., 1985, Radiative processes in astrophysics. John Wiley & Sons

SKAO Science Team 2015, Technical Report SKA-SCI-LOW-001, SKA1-Low Configuration. SKA Organisation

Sadoun R., Zheng Z., Miralda-Escudé J., 2017, ApJ, 839, 44

Sakharov A. D., 1966, Soviet Journal of Experimental and Theoretical Physics, 22, 241

Santos M. R., 2004, MNRAS, 349, 1137

Santos M. G., Ferramacho L., Silva M. B., Amblard A., Cooray A., 2010, MNRAS, 406, 2421

Santos S., Sobral D., Matthee J., 2016, MNRAS, 463, 1678

Scannapieco E., Harlow F. H., 1995, Introduction to finite-difference methods for numerical fluid dynamics, doi:10.2172/212567.

Schaal K., Bauer A., Chandrashekar P., Pakmor R., Klingenberg C., Springel V., 2015, MNRAS, 453, 4278

Schaye J., 2001, ApJ, 559, 507

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

Schmidt M., 1959, ApJ, 129, 243

Schmidt M., 1963, Nature, 197, 1040

Shapiro P. R., 1986, PASP, 98, 1014

Shapiro P. R., Giroux M. L., 1987, ApJ, 321, L107

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65

Sharma M., Theuns T., Frenk C., Bower R., Crain R., Schaller M., Schaye J., 2016, MNRAS, 458, L94

Sheth R. K., Mo H. J., Tormen G., 2001, MNRAS, 323, 1

Shibuya T., et al., 2014, ApJ, 788, 74

Shibuya T., et al., 2018, PASJ, 70, S14

Shimabukuro H., Semelin B., 2017, MNRAS, 468, 3869

Silva M. B., Kooistra R., Zaroubi S., 2016, MNRAS, 462, 1961

Sims P. H., Lentati L., Alexander P., Carilli C. L., 2016, MNRAS, 462, 3069

Slipher V. M., 1917, Proceedings of the American Philosophical Society, 56, 403

Smit R., et al., 2018, Nature, 553, 178

Smith R. E., Sheth R. K., Scoccimarro R., 2008, Phys. Rev. D, 78, 023523

Smith A., Ma X., Bromm V., Finkelstein S. L., Hopkins P. F., Faucher- Giguère C.-A., Kereš D., 2018, preprint, p. arXiv:1810.08185 (arXiv:1810.08185)

Sobacchi E., Mesinger A., 2015, MNRAS, 453, 1843

Sobacchi E., Mesinger A., Greig B., 2016, MNRAS, 459, 2741

Sobral D., Matthee J., Darvish B., Schaerer D., Mobasher B., Röttgering H. J. A., Santos S., Hemmati S., 2015, ApJ, 808, 139

Sobral D., et al., 2017, preprint, p. arXiv:1710.08422 (arXiv:1710.08422)

Songaila A., Hu E. M., Barger A. J., Cowie L. L., Hasinger G., Rosenwasser B., Waters C., 2018, ApJ, 859, 91

Spergel D. N., et al., 2003, ApJS, 148, 175

Spergel D., et al., 2015, arXiv e-prints, p. arXiv:1503.03757

Springel V., 2005, Monthly Notices of the Royal Astronomical Society, 364, 1105

Springel V., 2010a, ARA&A, 48, 391

Springel V., 2010b, MNRAS, 401, 791

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., Yoshida N., White S. D. M., 2001, New A, 6, 79

Springel V., et al., 2005, Nature, 435, 629

Stark D. P., Ellis R. S., Chiu K., Ouchi M., Bunker A., 2010, MNRAS, 408, 1628

Stark D. P., Ellis R. S., Ouchi M., 2011, ApJ, 728, L2

Stark D. P., et al., 2015, MNRAS, 450, 1846

Stark D. P., et al., 2017, MNRAS, 464, 469

Starobinsky A. A., 1980, Physics Letters B, 91, 99

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, ApJ, 717, 289

Steigman G., 2007, Annual Review of Nuclear and Particle Science, 57, 463

Strömberg G., 1925, Proceedings of the National Academy of Science, 11, 365

Strömgren B., 1932, ZAp, 4, 118

Takada M., et al., 2014, PASJ, 66, R1

Tamura N., et al., 2018, in Ground-based and Airborne Instrumentation for Astronomy VII. p. 107021C, doi:10.1117/12.2311871

Tauscher K., Rapetti D., Burns J. O., 2020, arXiv e-prints, p. arXiv:2005.00034

Tegmark M., Zaldarriaga M., 2002, Phys. Rev. D, 66, 103508

Tegmark M., Silk J., Rees M. J., Blanchard A., Abel T., Palla F., 1997, ApJ, 474, 1

Tepper-García T., 2006, MNRAS, 369, 2025

Teyssier R., 2002, A&A, 385, 337

Teyssier R., 2015, ARA&A, 53, 325

Tingay S. J., et al., 2013, PASA, 30, e007

Toro E., 1997, Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer, doi:10.1007/b79761

Totani T., Kawai N., Kosugi G., Aoki K., Yamada T., Iye M., Ohta K., Hattori T., 2006, PASJ, 58, 485

Totani T., Aoki K., Hattori T., Kawai N., 2016, PASJ, 68, 15

Tozzi P., Madau P., Meiksin A., Rees M. J., 2000, ApJ, 528, 597

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

Trenti M., Stiavelli M., 2009, ApJ, 694, 879

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., Trenti M., Stiavelli M., Auger M. W., Bradley L. D., 2012, ApJ, 747, 27

Trombetti T., Burigana C., 2018, Frontiers in Astronomy and Space Sciences, 5, 33

Troxel M. A., et al., 2018, Phys. Rev. D, 98, 043528

Tumlinson J., 2000, in Weiss A., Abel T. G., Hill V., eds, The First Stars. p. 185, doi:10.1007/10719504_32

Verde L., et al., 2002, MNRAS, 335, 432

Verhamme A., Schaerer D., Maselli A., 2006, A&A, 460, 397

Verhamme A., et al., 2018, MNRAS, 478, L60

Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, Nature Reviews Physics, 2, 42

Vrbanec D., et al., 2016, MNRAS, 457, 666

Walker A. G., 1935, MNRAS, 95, 263

Waterson M. F., Labate M. G., Schnetler H., Wagg J., Turner W., Dewdney P., 2016, in Proc. SPIE. p. 990628, doi:10.1117/12.2232526

Wayth R. B., et al., 2018, PASA, 35, 33

Weinberg S., 1972, Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity

Weinberg S., 2008, Cosmology

Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2564

Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wiersma R. P. C., et al., 2013, MNRAS, 432, 2615

Wild J. P., 1952, ApJ, 115, 206

Wilkes M., 1956, Automatic Digital Computaters. Methuen

Witstok J., Puchwein E., Kulkarni G., Smit R., Haehnelt M. G., 2019, arXiv e-prints, p. arXiv:1905.06954

Wood K., Loeb A., 2000, ApJ, 545, 86

Worseck G., et al., 2014, MNRAS, 445, 1745

Wyithe J. S. B., Bolton J. S., 2011, MNRAS, 412, 1926

Wyithe J. S. B., Morales M. F., 2007, MNRAS, 379, 1647

Xu G., 1995, ApJS, 98, 355

Yang H., Malhotra S., Gronke M., Rhoads J. E., Dijkstra M., Jaskot A., Zheng Z., Wang J., 2016, ApJ, 820, 130

Yoshiura S., Line J. L. B., Kubota K., Hasegawa K., Takahashi K., 2018, MNRAS, 479, 2767

Yoshiura S., Takahashi K., Takahashi T., 2020, Phys. Rev. D, 101, 083520

Zackrisson E., et al., 2020, MNRAS, 493, 855

Zahn O., Lidz A., McQuinn M., Dutta S., Hernquist L., Zaldarriaga M., Furlanetto S. R., 2007, ApJ, 654, 12

Zahn O., Mesinger A., McQuinn M., Trac H., Cen R., Hernquist L. E., 2011, MNRAS, 414, 727

Zeldovich Y. B., 1972, MNRAS, 160, 1P

Zeldovich Y. B., Sunyaev R. A., 1969, Ap&SS, 4, 301

Zel'dovich Y. B., Kurt V. G., Syunyaev R. A., 1969, Soviet Journal of Experimental and Theoretical Physics, 28, 146

Zheng Z., Wallace J., 2014, ApJ, 794, 116

Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2010, ApJ, 716, 574

Zheng Z.-Y., et al., 2017, ApJ, 842, L22

Zitrin A., et al., 2015, ApJ, 810, L12

van Haarlem M. P., et al., 2013, A&A, 556, A2