

# A new model framework for circumgalactic Ly $\alpha$ radiative transfer constrained by galaxy-Ly $\alpha$ forest clustering

Koki Kakiichi<sup>1\*</sup> and Mark Dijkstra<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, University College London, London WC1E 6BT, UK*

<sup>2</sup>*Institute of Theoretical Astrophysics, University of Oslo, Postboks 1029 Blindern, N-0315 Oslo, Norway*

Accepted 2018 August 6. Received 2018 July 18; in original form 2017 June 14

## ABSTRACT

We present a new perturbative approach to ‘constrained Ly  $\alpha$  radiative transfer’ (RT) through the circum- and intergalactic medium (CGM and IGM). We constrain the H I content and kinematics of the CGM and IGM in a physically motivated model, using the galaxy-Ly  $\alpha$  forest clustering data from spectroscopic galaxy surveys in quasar fields at  $z \sim 2-3$ . This enables us to quantify the impact of the CGM/IGM on Ly  $\alpha$  emission in an observationally constrained, realistic cosmological environment. Our model predicts that the CGM and IGM at these redshifts transmit  $\approx 80$  per cent of Ly  $\alpha$  photons after having escaped from galaxies. This implies that while the interstellar medium primarily regulates Ly  $\alpha$  escape, the CGM has a non-negligible impact on the observed Ly  $\alpha$  line properties and the inferred Ly  $\alpha$  escape fraction, even at  $z \sim 2-3$ . Ly  $\alpha$  scattering in the CGM and IGM further introduces an environmental dependence in the (apparent) Ly  $\alpha$  escape fraction, and the observed population of Ly  $\alpha$  emitting galaxies: the CGM/IGM more strongly suppresses direct Ly  $\alpha$  emission from galaxies in overdense regions in the Universe, and redistributes this emission into brighter Ly  $\alpha$  haloes. The resulting mean surface brightness profile of the Ly  $\alpha$  haloes is generally found to be a power law  $\propto r^{-2.4}$ . Although our model still contains arbitrariness, our results demonstrate how (integral field) spectroscopic surveys of galaxies in QSO fields constrain circumgalactic Ly  $\alpha$  RT, and we discuss the potential of these models for studying CGM physics and cosmology.

**Key words:** line: formation – radiative transfer – galaxies: high-redshift – intergalactic medium – quasars: absorption lines.

## 1 INTRODUCTION

Understanding the interaction between galaxies and the surrounding circum- and intergalactic medium (CGM and IGM) is key to understanding many outstanding problems in modern extragalactic astrophysics. The role of feedback and environment such as cold streams (Kereš et al. 2005; Dekel et al. 2009) and galactic winds (Murray, Quataert & Thompson 2005; Veilleux, Cecil & Bland-Hawthorn 2005) has been a central theme in regulating the evolution of galaxies throughout cosmic history (White & Frenk 1991; Croton et al. 2006; Hopkins et al. 2014; Schaye et al. 2015). The structure of the CGM and IGM around galaxies is influenced both by supernova and black hole feedback, as well as the large-scale gaseous environment. This makes the CGM an ideal laboratory to test galaxy formation theories.

The structure and kinematics of H I gas in the CGM and IGM directly affects observable properties of Ly  $\alpha$  line emission. Exam-

ples of these properties include the observed Ly  $\alpha$  spectral line shape (e.g. Gronke et al. 2016, 2017), and the surface brightness profiles of extended Ly  $\alpha$  haloes around galaxies (e.g. Cantalupo et al. 2014; Hennawi et al. 2015). A proper understanding of circumgalactic and intergalactic Ly  $\alpha$  radiative transfer (RT) would allow us to use observations of the Ly  $\alpha$  emission line into a unique window on the circum-/intergalactic gaseous environment.

In addition, understanding circum-/intergalactic Ly  $\alpha$  RT has several cosmological applications: (i) understanding how the IGM and CGM affect Ly  $\alpha$  line emission is crucial when using galaxies at  $z > 6$  to probe the Epoch of Reionization (e.g. Dijkstra, Mesinger & Wyithe 2011; Choudhury et al. 2015; Mesinger et al. 2015; Kakiichi et al. 2016; Mason et al. 2017; Sadoun, Zheng & Miralda-Escudé 2017, see e.g. Dijkstra 2014 for an extensive review) and (ii) cosmological surveys of Ly  $\alpha$  emitting galaxies  $z \sim 2-4$  which aim to constrain the dark energy equation of state (e.g. with Hobby-Eberly Telescope Dark Energy Experiment (HETDEX), Hill et al. 2008) may be sensitive to Ly  $\alpha$  RT effects, as the Ly  $\alpha$  RT through the IGM modulates the apparent visibility of Ly  $\alpha$  emission line of galaxies (Zheng et al. 2011, Behrens & Niemeyer

\* E-mail: k.kakiichi@ucl.ac.uk

2013, and this paper). The Ly $\alpha$  RT effect in the clustering of Ly $\alpha$  emission has tentatively been detected in the Baryon Oscillation Spectroscopic Survey (BOSS) data by Croft et al. (2016), although controversy still remains. Ly $\alpha$  RT effects can therefore affect redshift-space distortions (RSD) in way that can affect the accuracy with which cosmological parameters can be constrained (Wyithe & Dijkstra 2011). A combined analysis of the three-point correlation function would be required to circumvent the issue (Greig, Komatsu & Wyithe 2013).

On the theoretical side, modelling Ly $\alpha$  transfer through multi-phase interstellar medium (ISM), CGM and IGM from first principles is a daunting task (e.g. Hayes 2015, Gronke et al. 2016, 2017, McCourt et al. 2016, Dijkstra 2017), partially because existing numerical simulations do not have the resolution to properly resolve the cold, neutral phase (McCourt et al. 2016). In addition, it has been difficult to observationally disentangle the impact of the ISM, CGM and IGM on observed Ly $\alpha$  line shapes of galaxies: we currently do not understand to what extent observed Ly $\alpha$  spectral line shapes, and even observationally inferred Ly $\alpha$  escape fractions, are shaped by small-scale RT in the ISM inside galaxies versus the intermediate-to-large-scale CGM and IGM outside galaxies. Disentangling the two effects is of great interest for understanding the gaseous environment of galaxies (and therefore galaxy formation), and for being able to use Ly $\alpha$  emitting galaxy surveys for cosmological purposes.

Spectroscopic surveys of galaxies in the foregrounds of quasars (QSOs) provide a potentially unique observational window on Ly $\alpha$  RT on a range of scales, as they enable us to study both the ISM properties of galaxies from direct spectroscopy, and the surrounding CGM and IGM from the Ly $\alpha$  absorption in the background QSO spectra (at  $z \sim 2-3$ , see Adelberger et al. 2003, 2005; Cooke et al. 2006; Crighton et al. 2011; Rakic et al. 2012; Rudie et al. 2012; Tummuangpak et al. 2014; Turner et al. 2014). Despite the enormous scientific potential of this approach (Steidel et al. 2010, 2011; Hennawi & Prochaska 2013; Hennawi et al. 2015), a theoretical framework that is needed to fully exploit these observations is still in its infancy. Most previous works investigating the connection between galaxies/quasars and their gaseous environments focused either on Ly $\alpha$  emission (Cantalupo et al. 2005; Faucher-Giguère et al. 2010; Kollmeier et al. 2010; Lake et al. 2015; Mas-Ribas & Dijkstra 2016) or Ly $\alpha$  absorption around galaxies and quasars (Fumagalli et al. 2011; Shen et al. 2013; Meiksin, Bolton & Tittley 2015; Rahmati et al. 2015; Faucher-Giguère et al. 2016). Dijkstra & Kramer (2012) introduced ‘constrained Ly $\alpha$  RT’ in which Ly $\alpha$  circumgalactic RT modelling was performed through models of the CGM in which its H I content was constrained by the Ly $\alpha$  absorption signal in the background galaxies and QSOs (see also Noterdaeme et al. 2012; Krogager et al. 2013; Cai et al. 2017). This approach enables controlled numerical experiments of Ly $\alpha$  RT through physical realistic distributions of H I in the CGM/IGM.

In this paper, we present an improved constrained Ly $\alpha$  RT model of the impact of the CGM and IGM on Ly $\alpha$  emission in and around galaxies. Earlier work on circumgalactic/intergalactic Ly $\alpha$  RT either employed analytic models for the density and kinematics of H I gas in the CGM/IGM (Dijkstra, Lidz & Wyithe 2007, also see Santos 2004) or cosmological (hydro)simulations (Iliev et al. 2008; Zheng et al. 2010; Dayal, Maselli & Ferrara 2011; Laursen, Sommer-Larsen & Razoumov 2011), which – as mentioned above – currently do not resolve the cold, neutral phase of the CGM/IGM (see McCourt et al. 2016 for extended discussion). Dijkstra & Kramer (2012) and Hennawi & Prochaska (2013) employed a joint Ly $\alpha$  emission – absorption approach to circumgalactic Ly $\alpha$  RT. Their

models, none the less, paid attention mostly to the neutral gas *distribution* in the CGM and Ly $\alpha$  haloes; the absorption data were not used optimally to constrain the gas *kinematics*. This is important as gas kinematics is well known to play an important role in Ly $\alpha$  transfer process. Our model improves upon previous works in two important aspects: (i) the distribution and kinematics of neutral gas around galaxies are constrained by observations of the galaxy-Ly $\alpha$  forest clustering and its RSD, which provide direct observational constraints on both the H I distribution and kinematics in the CGM/IGM; and (ii) we model both the impact of the CGM/IGM on the emerging Ly $\alpha$  flux from galaxies (and its spectral line shape), and the predict surface brightness profiles of the scattered Ly $\alpha$  radiation. The latter is important as Ly $\alpha$  scattering simultaneously attenuates direct Ly $\alpha$  emission from galaxies, and gives rise to spatially extended Ly $\alpha$  haloes. Thus, the two observables combined provide a more complete picture of how the CGM and IGM affect Ly $\alpha$  radiation in and around galaxies.

Our aim of this paper is twofold. The *first* is to introduce a new unified statistical, perturbative modelling of Ly $\alpha$  RT and galaxy-Ly $\alpha$  forest clustering in order to achieve the goals (i) and (ii). Our model naturally includes absorbers with H I column densities in the range  $\log_{10} N_{\text{H I}}/\text{cm}^{-2} \sim 13 - 22$ . This is important, as the nature of self-shielding absorbers is still not well understood. Moreover, the impact of low column density absorbers surrounding galaxies on the emerging Ly $\alpha$  emission line has not been explored at all yet. The *second* is to demonstrate how joint Ly $\alpha$  emission-absorption data can be used to obtain new insights into how Ly $\alpha$  radiation escapes from galaxies, how it subsequently propagates through the CGM/IGM, and how we can use this knowledge to obtain constraints on the physical properties of the CGM/IGM.

This paper is organized as follows. In Section 2, we summarize the methodology and the general idea behind our work. In Section 3, we introduce a unified approach to cosmological Ly $\alpha$  RT connecting Ly $\alpha$  emission and absorption. In Section 4, we describe a physical model of galaxies and the gaseous environments used to analyse the data. The joint analysis of the redshift-space anisotropic galaxy-Ly $\alpha$  forest clustering with the Ly $\alpha$  line profiles and Ly $\alpha$  haloes is presented in Section 5. In Section 6, we discuss the implications for the physics of Ly $\alpha$  escape and Ly $\alpha$  haloes in the CGM and cosmology. Our conclusions are summarized in Section 7. Throughout this paper, we adopt the flat quasars  $\Lambda$  cold dark matter ( $\Lambda$ CDM) cosmology with  $\Omega_{\text{m}} = 0.3$ ,  $\Omega_{\Lambda} = 0.7$ , and  $h = 0.7$ . We denote pkpc and pMpc (ckpc and cMpc) to indicate distances in proper (comoving) units.

## 2 METHODOLOGY: CONSTRAINED RADIATIVE TRANSFER

In this section, we describe how our general methodology – ‘constrained RT’ – works. A goal is to estimate the impact of the CGM and IGM around galaxies on the Ly $\alpha$  escape fraction and Ly $\alpha$  haloes consistently with the galaxy-Ly $\alpha$  forest clustering measurements (Section 5). We will discuss the result in the context of the physics of the galaxy-IGM connection and cosmology (Section 6).

Fig. 1 schematically illustrates how the joint analysis of Ly $\alpha$  emission and absorption from a (integral field) spectroscopic survey of galaxies in QSO fields can be used to study the origin of escape fraction and Ly $\alpha$  haloes. This is the most important picture which we base our argument on so that it should be kept in mind throughout the paper.

First, using galaxy-QSO pairs (black arrow), the galaxy-Ly $\alpha$  forest cross-correlation and its RSD can be measured. This can



respect to the number of scatterings. Ly $\alpha$  photons experiencing zero or single scattering contribute to the largest to the emergent Ly $\alpha$  profile. In the *zeroth-order scattering expansion*, we include the Ly $\alpha$  photons that have experienced no scattering; once a photon is scattered out of a line of sight, the photon is lost. Thus, Ly $\alpha$  photons are attenuated by  $e^{-\tau_{\text{CGM/IGM}}}$ . The zeroth-order scattering includes the Ly $\alpha$  absorption in the background QSOs and the CGM/IGM attenuation effect on the galaxy spectra (see Fig. 1). In the *first-order scattering expansion*, we include the contribution from Ly $\alpha$  photons that have scattered back into the line of sight after the first scattering. This allows us to estimate the Ly $\alpha$  haloes (see Fig. 1). In the higher order scattering expansion, we can extend this to include the contributions from the photons that experienced multiple scatterings, which will approach full Monte Carlo Ly $\alpha$  RT simulations.

Therefore, the perturbative approach enables us to self-consistently formulate the Ly $\alpha$  transfer problem both in absorption and emission, connecting the galaxy-Ly $\alpha$  forest clustering (Section 3.1) with the Ly $\alpha$  emission-line profiles of galaxy spectra (*the zeroth-order scattering expansion*) (Section 3.2) and Ly $\alpha$  haloes around galaxies (*the first-order scattering expansion*) (Section 3.3).

### 3.1 Ly $\alpha$ absorption around galaxies

The Ly $\alpha$  absorption arises from both residual neutral gas in the photoionized IGM and self-shielded gas. The Ly $\alpha$  absorption features in QSO spectra are a natural consequence of the continuous cosmic web of the large-scale structure in  $\Lambda$ CDM cosmology (e.g. Cen et al. 1994; Lukić et al. 2015). In this paper, in order to make the modelling analytically tractable, we approximate the CGM and IGM as consisting of a collection of Lagrangian gas parcels which are referred to as absorbers.

We define each absorber's velocity and position relative to a galaxy of interest. An absorber located at a comoving distance  $r$  away from the associate galaxy has its own H I column density  $N_{\text{HI}}$ , temperature  $T$ , and a proper peculiar velocity  $v_r$  relative to the galaxy (see Fig. 1). The total line-of-sight velocity of an absorber relative to the associated galaxy is  $u_{\parallel} = H(z)r_{\parallel}/(1+z) + v_{\parallel}$  where  $v_{\parallel} = \mu v_r$  is the line-of-sight peculiar velocity,  $\mu = r_{\parallel}/r$ , and  $r_{\parallel}$  is the comoving distance along the line of sight.  $H(z)$  is the Hubble parameter at redshift  $z$ . Thus, for the photons emitted at frequency  $\nu_e$  from a galaxy or background QSO, the Ly $\alpha$  optical depth of an absorber is given by

$$\tau_a(\nu_e | u_{\parallel}, N_{\text{HI}}) = \sigma_{\alpha} N_{\text{HI}} \varphi_{\nu} \left[ T, \nu_e \left( 1 - \frac{u_{\parallel}}{c} \right) \right], \quad (1)$$

where  $\sigma_{\alpha} = (\pi e^2 / m_e c) f_{12} = 0.011 \text{ cm}^2 \text{ Hz}$  is the Ly $\alpha$  cross-section,  $\varphi_{\nu}(T, \nu)$  is the Voigt profile at gas temperature  $T$ , frequency  $\nu$ ,  $e$  is the electron charge,  $m_e$  is the mass of electron,  $c$  is the speed of light, and  $f_{12} = 0.4164$  is the oscillator strength of  $2P \rightarrow 1S$  transition (Wiese & Fuhr 2009). In the following sections, we interchangeably express the emitted frequency  $\nu_e$  at the rest frame of galaxies in terms of the velocity unit  $\Delta v$  defined as  $\nu_e = \nu_{\alpha}(1 - \Delta v/c)$  and the proper line-of-sight redshift-space coordinate  $s_{\parallel}$  defined as  $s_{\parallel} = \Delta v/H(z)$  (in pMpc) where  $\nu_{\alpha}$  is the frequency at the Ly $\alpha$  line centre.

#### 3.1.1 The redshift-space anisotropic galaxy-Ly $\alpha$ forest clustering

The cross-correlation of the transmitted Ly $\alpha$  flux in the Ly $\alpha$  forest region of the background QSO spectra with galaxies' positions traces the statistical properties of the gaseous environments around

galaxies. The 2D effective optical depth map,  $\tau_{\text{eff}}(\nu_e, r_{\perp})$ , is defined in terms of the mean transmitted Ly $\alpha$  forest flux around galaxies,  $F(\nu_e, r_{\perp})$ , as a function of emitted frequency,  $\nu_e$ , at the rest frame of the galaxies and impact parameter,  $r_{\perp}$ , relative to the background QSO (the corresponding proper redshift-space coordinates are  $s_{\parallel} = (c/H)(1 - \nu_e/\nu_{\alpha})$  and  $s_{\perp} = r_{\perp}/(1+z)$ ):

$$\tau_{\text{eff}}(\nu_e, r_{\perp}) \equiv -\ln \langle F(\nu_e, r_{\perp}) \rangle. \quad (2)$$

As a solution to the statistical perturbative approach to cosmological Ly $\alpha$  transfer (Kakiichi et al. 2016, see also Paresce et al. 1980; Haardt & Madau 1996), we find that the 2D effective optical depth map is given by

$$\tau_{\text{eff}}(\nu_e, r_{\perp}) = \int dN_{\text{HI}} \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \left| \frac{dz}{dr} \right| \times \int \frac{du_{\parallel}}{aH} [1 + \xi_v(u_{\parallel}, r_{\perp})] [1 - e^{-\tau_a(\nu_e | u_{\parallel}, N_{\text{HI}})}], \quad (3)$$

where  $\frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z}$  is the H I column density distribution function (CDDF) of absorbers,  $|dr/dz| = c/H(z)$  is the comoving distance per redshift. The velocity-space correlation function between galaxies and absorbers,  $\xi_v(u_{\parallel}, r_{\perp})$ , in the Gaussian streaming model<sup>2</sup> is given by

$$1 + \xi_v(u_{\parallel}, r_{\perp}) = \int_{-\infty}^{\infty} \frac{aH dr_{\parallel}}{\sqrt{2\pi\sigma_v^2(r)}} [1 + \xi(r)] \exp \left[ -\frac{(u_{\parallel} - aHr_{\parallel} - \mu\langle v_r(r) \rangle)^2}{2\sigma_v^2(r)} \right], \quad (4)$$

where  $\xi(r)$  is the real-space correlation function between galaxies and absorbers,  $\langle v_r(r) \rangle$  and  $\sigma_v(r)$  are the mean radial velocity and velocity dispersion between galaxies and absorbers at a separation  $r$ ,  $r = \sqrt{r_{\parallel}^2 + r_{\perp}^2}$ , and  $a = 1/(1+z)$ . For simplicity, we assume isotropic velocity dispersion so that both line-of-sight and radial components are the same.

The velocity-space correlation function takes into account the dynamics and clustering of absorbers around galaxies. While we use the Gaussian streaming model as an explicit example, the method can easily be generalized for any pairwise velocity probability distribution function. We model the explicit forms of real-space correlation function and the pairwise mean velocity and velocity dispersion in Section 4.

#### 3.1.2 The equivalent width of Ly $\alpha$ absorption

The rest-frame equivalent width (EW) is computed using the same method as Rakic et al. (2012), which measures the excess absorption relative to the mean IGM. The mean EW of Ly $\alpha$  absorption around galaxies is then computed as

$$\langle \text{EW}(r_{\perp}) \rangle = \lambda_{\alpha} \int \frac{d\nu_e}{\nu_{\alpha}} \left[ 1 - \frac{\langle F(\nu_e, r_{\perp}) \rangle}{\bar{F}(z)} \right], \quad (5)$$

where  $\bar{F}(z) = e^{-\bar{\tau}_{\text{eff}}(z)}$  and  $\bar{\tau}_{\text{eff}}(z)$  is the mean effective optical depth of Becker et al. (2013). In the case of  $\xi_v = 0$ , equation (3) reproduces the mean effective optical depth.

<sup>2</sup>Fisher (1995) and Reid & White (2011) have shown that the Gaussian streaming model gives the same result as linear theory including Kaiser effect provided that the mean radial velocity and velocity dispersion are computed from linear theory. Because we are interested in the gas flow around galaxies on scales of a few Mpc we do not impose the linear theory. Instead we allow a phenomenological parametrization of the velocity field.



### 3.2 Ly $\alpha$ emission-line profiles of galaxies

The Ly  $\alpha$  emission-line profiles of galaxies are affected by both the ISM-scale RT and the large-scale CGM/IGM environment. We therefore split the model into the small-scale RT part (Section 3.2.1), which includes the multiple scatterings, and the large-scale RT part (Section 3.2.2), which is modelled self-consistently to the Ly  $\alpha$  absorption and diffuse Ly  $\alpha$  haloes around galaxies using the perturbative expansion.

#### 3.2.1 Star formation and ISM of galaxies

For Ly  $\alpha$  emission due to the nebular recombination in star-forming regions, the average Ly  $\alpha$  luminosity is given by

$$\langle L_{\alpha}^{\text{intr}} \rangle = 1.1 \times 10^{42} \left( \langle \text{SFR} \rangle / M_{\odot} \text{ yr}^{-1} \right) \text{ erg s}^{-1}, \quad (6)$$

by converting the relation of H  $\alpha$  luminosity and star formation rate (SFR, Kennicutt 1998) using the case B approximation (Brocklehurst 1971). We assume the average SFR of galaxies is  $\langle \text{SFR} \rangle = 34 M_{\odot} \text{ yr}^{-1}$ , which is taken from the median dust-corrected SFR of Lyman-Break Galaxies (LBGs) in Steidel et al. (2011). Therefore, our fiducial value is  $\langle L_{\alpha}^{\text{intr}} \rangle = 3.7 \times 10^{43} \text{ erg s}^{-1}$ .

We assume dust-free multiphase clumpy ISM. We define the ISM Ly  $\alpha$  escape fraction,  $\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle$ , as a fraction of Ly  $\alpha$  photons escaped from the sites of nebular Ly  $\alpha$  emission out of galaxies and toward the line of sight of an observer. We assume a fiducial value of  $\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle = 0.20$ . Note that in a dust-free ISM only the Ly  $\alpha$  scattering by the H I gas contributes to the ISM escape fraction by scattering off Ly  $\alpha$  photons out of the line of sight of an observer. Thus in the absence of the CGM and IGM, the observed Ly  $\alpha$  luminosity is  $\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle \langle L_{\alpha}^{\text{intr}} \rangle$ . However, because these scattered photons eventually leak out of the ISM, the photons escaping out of the ISM in *all directions* around a galaxy are the same as the intrinsic nebular production of Ly  $\alpha$  photons. In other words, the amount of Ly  $\alpha$  photons injected into the CGM is the same as  $\langle L_{\alpha}^{\text{intr}} \rangle$  in a dust-free ISM model.

For the intrinsic average Ly  $\alpha$  line profile,  $\langle \Phi_{\alpha}^{\text{ISM}}(v_e) \rangle$ , (i.e. emerging from the ISM), we use a result from the Monte Carlo Ly  $\alpha$  RT model of the multiphase clumpy ISM of Gronke & Dijkstra (2016) with the H I column density of  $\log_{10}(N_{\text{H I}}/\text{cm}^{-2}) = 19.64$  and outflow velocity of  $202.2 \text{ km s}^{-1}$  (one presented in fig. 6 of Gronke & Dijkstra 2016). This model produces an asymmetric single peak line profile with the Ly  $\alpha$  velocity offset of  $\langle \Delta v_{\text{Ly}\alpha} \rangle = 250 \text{ km s}^{-1}$ .

Both the ionizing and non-ionizing UV luminosity incorporate results from stellar population synthesis. The average Lyman continuum (LyC,  $< 912 \text{ \AA}$ ) photon production rate (in units of  $\text{s}^{-1}$ ) is given by assuming the following conversion from the ultraviolet (UV,  $1500 \text{ \AA}$ ) luminosity (Robertson et al. 2013),

$$\begin{aligned} \langle \dot{N}_{\text{ion}} \rangle &= \langle f_{\text{esc}}^{\text{LyC}} \rangle \langle \xi_{\text{ion}} \rangle \langle L_{\text{UV}} \rangle, \\ &\approx 8.6 \times 10^{52} \left( \frac{\langle f_{\text{esc}}^{\text{LyC}} \rangle}{0.02} \right) \left( \frac{\langle \text{SFR} \rangle}{34 M_{\odot} \text{ yr}^{-1}} \right) \text{ s}^{-1}, \end{aligned} \quad (7)$$

where we used

$$\langle L_{\text{UV}} \rangle = 8 \times 10^{27} \left( \langle \text{SFR} \rangle / M_{\odot} \text{ yr}^{-1} \right) \text{ erg s}^{-1} \text{ Hz}^{-1}, \quad (8)$$

with the assumptions of a Salpeter initial mass function and solar metallicity (Madau, Pozzetti & Dickinson 1998), and  $\log_{10}(\xi_{\text{ion}})/\text{erg Hz}^{-1} = 25.2$ , consistent with stellar population synthesis model (Robertson et al. 2013) and observations (Bouwens et al. 2016). We assume an average LyC escape fraction of  $\langle f_{\text{esc}}^{\text{LyC}} \rangle = 0.02$ , which is broadly in agreement with the observationally inferred values of the LyC escape fractions (e.g. Siana et al.

2015; Grazian et al. 2016) at  $z \sim 3$ . In this model, the average intrinsic EW of Ly  $\alpha$  emission line is  $\langle W_{\text{Ly}\alpha}^{\text{intr}} \rangle = 103 \text{ \AA}$ , assuming the UV slope of  $-2$  at  $z \sim 3$  (Bouwens et al. 2014).

#### 3.2.2 Emergent Ly $\alpha$ line profile from CGM/IGM

Ly  $\alpha$  photons escaping from the ISM are scattered out of a line of sight by the intervening CGM and IGM. Thus, the emergent average Ly  $\alpha$  line profile of Ly  $\alpha$  emitting galaxies is affected by the opacity of the circum- and intergalactic gaseous environments. This effect has been formulated by Kakiichi et al. (2016) and the CGM/IGM attenuates the average Ly  $\alpha$  flux of galaxies as

$$\langle L_{\nu}(v_e) \rangle = e^{-\tau_{\text{eff}}^{\text{Ly}\alpha}(v_e)} \left\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \right\rangle \langle L_{\alpha}^{\text{intr}} \rangle \langle \Phi_{\alpha}^{\text{ISM}}(v_e) \rangle, \quad (9)$$

where  $\langle L_{\nu}(v_e) \rangle$  is the apparent specific Ly  $\alpha$  luminosity seen by an observer and  $\tau_{\text{eff}}^{\text{Ly}\alpha}(v_e)$  is effective optical depth against Ly  $\alpha$  line at the emitted frequency  $v_e$  in the rest frame of galaxies. The mean transmissivity of the Ly  $\alpha$  flux of galaxies is defined as

$$\langle \mathcal{T}_{\alpha} \rangle = \int e^{-\tau_{\text{eff}}^{\text{Ly}\alpha}(v_e)} \langle \Phi_{\alpha}^{\text{ISM}}(v_e) \rangle dv_e. \quad (10)$$

The observed Ly  $\alpha$  EW of the emission line is then given by  $\langle W_{\text{Ly}\alpha} \rangle = \langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle \langle \mathcal{T}_{\alpha} \rangle \langle W_{\text{Ly}\alpha}^{\text{intr}} \rangle$ .

The CGM/IGM attenuation is caused by the same neutral gas that gives rise to Ly  $\alpha$  absorption around galaxies observed in the background QSO spectra. Therefore, our modelling leads the effective optical depth against Ly  $\alpha$  line along each sight line of galaxies as,

$$\begin{aligned} \tau_{\text{eff}}^{\text{Ly}\alpha}(v_e) &= \int dN_{\text{H I}} \frac{\partial^2 \mathcal{N}}{\partial N_{\text{H I}} \partial z} \left| \frac{dz}{dr} \right| \\ &\times \int \frac{du_{\parallel}}{aH} [1 + \zeta_{\nu}(u_{\parallel})] [1 - e^{-\tau_{\alpha}(v_e|u_{\parallel}, N_{\text{H I}})}]. \end{aligned} \quad (11)$$

Clearly, the CGM/IGM attenuation of the Ly  $\alpha$  emission line of galaxies is closely related to the effective optical depth for the galaxy-Ly  $\alpha$  absorption clustering (equation 3). The difference is that the former is affected only by the foreground gas of galaxies, whereas the latter must also include the gas background of the galaxies. Therefore, the velocity–space correlation function between galaxies and absorbers,  $\zeta_{\nu}(u_{\parallel})$ , takes the integral from the innermost radius  $r_{\text{min}}$  away from a galaxy to an observer at infinity,

$$\begin{aligned} 1 + \zeta_{\nu}(u_{\parallel}) &= \int_{r_{\text{min}}}^{\infty} \frac{aH dr_{\parallel}}{\sqrt{2\pi\sigma_{\nu}^2(r_{\parallel})}} \\ &\times \left[ 1 + \xi(r_{\parallel}) \right] \exp \left[ -\frac{(u_{\parallel} - aHr_{\parallel} - \langle v_r(r_{\parallel}) \rangle)^2}{2\sigma_{\nu}^2(r_{\parallel})} \right]. \end{aligned} \quad (12)$$

The real-space correlation function  $\xi(r_{\parallel})$ , the mean radial velocity field  $\langle v_r(r_{\parallel}) \rangle$ , and velocity dispersion  $\sigma_{\nu}(r_{\parallel})$  are the same as ones in equation (4), but evaluated at  $r_{\parallel}$ . We adopt the value of  $r_{\text{min}} = 80 \text{ pkpc}$ , which corresponds to the virial radius of dark matter haloes of mass  $M_{\text{h}} \approx 9 \times 10^{11} M_{\odot}$  (Steidel et al. 2010). In Appendix C, we have tested the dependence of the Ly  $\alpha$  transmission against the choice of the innermost radius from  $r_{\text{min}} = 30$  to  $160 \text{ pkpc}$  bracketing the value ( $r_{\text{min}} = 1.5r_{\text{vir}}$ ) suggested by Laursen et al. (2011).

The higher order Lyman series scattering by the intervening circum-/intergalactic gas also contributes to the opacity against the UV radiation from galaxies at the rest-frame wavelength  $912 \text{ \AA} < \lambda < 1216 \text{ \AA}$ . The total effective optical depth by the line blanketing is  $e^{-\tau_{\text{eff}}^{\text{line}}(v_e)} = \sum_{n=\alpha, \beta, \dots} e^{-\tau_{\text{eff}}^{\text{Ly}n}(v_e)}$ . The calculation of the effective

optical depth against the higher order Ly $\alpha$  line,  $\tau_{\text{eff}}^{\text{Ly}\alpha}(\nu_e)$ , is identical to that of Ly $\alpha$  except for replacing the Ly $\alpha$  cross-section with the higher order Ly $\alpha$  cross-section.

### 3.2.3 Ly $\alpha$ escape fraction

We define the average total Ly $\alpha$  escape fraction,  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle$ , as a fraction of Ly $\alpha$  photons escaped from the sites of nebular Ly $\alpha$  emission within the ISM, through the CGM and IGM, and reached to a finite aperture of a telescope, averaged over many galaxies. Thus, the average total Ly $\alpha$  escape fraction is expressed as a product of the ISM Ly $\alpha$  escape fraction<sup>3</sup> and the Ly $\alpha$  transmissivity of the CGM and IGM,

$$\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle = \langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle \langle \mathcal{T}_{\alpha} \rangle. \quad (13)$$

## 3.3 Ly $\alpha$ haloes around galaxies

While the various formation mechanisms of diffuse Ly $\alpha$  haloes are proposed, the physical origin of Ly $\alpha$  haloes is still unknown. The sources of Ly $\alpha$  emission in the diffuse haloes can be generated by (1) scattered Ly $\alpha$  photons from the central galaxies, (2) fluorescent Ly $\alpha$  radiation by the ionizing photons from the galaxies, (3) cooling radiation, and (4) Ly $\alpha$  emission from the unresolved galaxies around the detected galaxies. In this paper, we focus on the Ly $\alpha$  scattering contribution to Ly $\alpha$  haloes.

Our perturbative approach can however be extended to include the contribution from the other Ly $\alpha$  emission such as fluorescence. We refer the reader to Mas-Ribas et al. (2017) for discussion on the observational test for distinguishing different origins of Ly $\alpha$  haloes.

### 3.3.1 Ly $\alpha$ emissivity from scattered radiation

In the first-order scattering expansion of cosmological Ly $\alpha$  transfer, we take into account the contribution from the photons that are scattered once. When Ly $\alpha$  haloes are formed due to the scattering of Ly $\alpha$  photons from the central sources, all the photons experienced no scattering (zeroth-order expansion) do not contribute to the Ly $\alpha$  haloes. This is equivalent to the *single scattering approximation*,<sup>4</sup> in which we truncate the perturbative expansion in the first order.

As Ly $\alpha$  photons are scattered by the neutral gas around galaxies, each absorber absorbs and re-emits the Ly $\alpha$  photons from the central Ly $\alpha$  emitting galaxy with specific luminosity  $L_{\nu}^{\text{intr}}(\nu_e) = \langle L_{\alpha}^{\text{intr}} \rangle \langle \Phi_{\alpha}^{\text{ISM}}(\nu_e) \rangle$ . Therefore, the surface brightness of Ly $\alpha$  haloes is given by integrating the Ly $\alpha$  emission from all the absorbers over the distribution and kinematics.

First, we derive the individual absorber's luminosity illuminated by the scattering of Ly $\alpha$  photons emitted from a central Ly $\alpha$  emitting galaxy. The probability that an absorber absorbs Ly $\alpha$  radiation is  $1 - e^{-\tau_{\alpha}}$ . Only the area extended by an absorber can receive the radiation emitted from the central galaxy, therefore the Ly $\alpha$  radiation received by an absorber is  $\sigma_{\text{abs}} L_{\nu}^{\text{intr}} / (4\pi l_p^2)$

<sup>3</sup>The ISM Ly $\alpha$  escape fraction is defined as a fraction of Ly $\alpha$  photons escaped from the sites of nebular Ly $\alpha$  emission within the ISM to outside of galaxies.

<sup>4</sup>Dijkstra & Kramer (2012) has employed the single scattering approximation, but formulated differently from this paper. They have compared the analytic Ly $\alpha$  RT model based on the single scattering approximation with the Monte Carlo Ly $\alpha$  RT simulations, and showed that the single scattering approximation works well.

where  $\sigma_{\text{abs}}$  is the geometrical cross section and  $l_p = r/(1+z)$  is the proper distance. The absorber absorbs the emitted photons of frequency  $\nu_e$  from the central galaxy by the amount of  $(1 - e^{-\tau_{\alpha}})\sigma_{\text{abs}} L_{\nu}^{\text{intr}}(\nu_e) / (4\pi l_p^2 h\nu_e)$  [photons s<sup>-1</sup>] and re-emits the photons near the Ly $\alpha$  line centre. Thus, the individual absorber's Ly $\alpha$  luminosity (in units of erg s<sup>-1</sup>) due to the scattering is given by,

$$L_{\alpha}^{\text{abs}}(r, v_r, N_{\text{HI}}) = \int [1 - e^{-\tau_{\alpha}(v_{\text{inj}}, N_{\text{HI}})}] \frac{\sigma_{\text{abs}} L_{\nu}^{\text{intr}}(\nu_e)}{4\pi l_p(r)^2} dv_e, \quad (14)$$

where  $v_{\text{inj}} = \nu_e [1 - (H(z)l_p(r) + v_r)/c]$  is the frequency injected to an absorber.

Then, these individual absorber's Ly $\alpha$  luminosities integrate overall H I column density and velocities of the absorber. The velocity of an absorber affects the probability that Ly $\alpha$  photons from the central sources are scattered through its velocity dependence in the optical depth. The bolometric Ly $\alpha$  emissivity  $\langle \epsilon_{\alpha}(r) \rangle$  (in units of erg s<sup>-1</sup> cMpc<sup>-3</sup>) from the absorbers around galaxies is given by (see Appendix A for derivation),

$$\begin{aligned} \langle \epsilon_{\alpha}(r) \rangle &= \frac{1 + \xi(r)}{4\pi r^2} \int dN_{\text{HI}} \left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\frac{dz}{dr}} \\ &\times \int \frac{dv_r}{\sqrt{2\pi\sigma_v^2(r)}} \mathcal{L}_{\alpha}(r, v_r, N_{\text{HI}}) \exp \left[ -\frac{(v_r - \langle v_r(r) \rangle)^2}{2\sigma_v^2(r)} \right] \end{aligned} \quad (15)$$

where  $\mathcal{L}_{\alpha}(r, v_r, N_{\text{HI}})$  is the Ly $\alpha$  luminosity of the scattered photons by the absorbers around the central galaxies of specific luminosity  $L_{\nu}$ :

$$\mathcal{L}_{\alpha}(r, v_r, N_{\text{HI}}) = \int [1 - e^{-\tau_{\alpha}(v_{\text{inj}}, N_{\text{HI}})}] L_{\nu}^{\text{intr}}(\nu_e) dv_e. \quad (16)$$

The calculation of the Ly $\alpha$  halo emissivity does not depend on the choice of the innermost radius  $r_{\text{min}}$ . The Ly $\alpha$  halo profile at small radii,  $r < r_{\text{min}}$ , can be self-consistently computed as long as we are concerned with the contribution from the singly scattered photons, which is a fundamental ansatz of the perturbative approach.

### 3.3.2 The surface brightness of Ly $\alpha$ haloes

The mean surface brightness,  $\langle \text{SB}_{\alpha}(r_{\perp}) \rangle$ , (in units of erg s<sup>-1</sup> cm<sup>-2</sup> sr<sup>-1</sup>) of Ly $\alpha$  haloes due to the scattered Ly $\alpha$  photons from the central galaxies at impact parameter  $r_{\perp}$  is then given by,

$$\langle \text{SB}_{\alpha}(r_{\perp}) \rangle = \frac{(1+z)^2}{4\pi(1+z)^4} \int_{r_{\perp}}^{\infty} \frac{2r dr}{\sqrt{r^2 - r_{\perp}^2}} \langle \epsilon_{\alpha}(r) \rangle. \quad (17)$$

We apply the factor of  $(1+z)^2$  to convert to the proper unit and the factor of  $(1+z)^{-4}$  to convert it to the surface brightness value observed at  $z=0$ .

We also introduce an Ly $\alpha$  halo flux fraction  $X_{\text{Ly}\alpha, \text{halo}}$  as defined by Wisotzki et al. (2016), which is a fraction of the integrated Ly $\alpha$  flux in a Ly $\alpha$  halo relative to the sum of the central galaxy's Ly $\alpha$  line flux  $F_{\text{gal}}$  and the Ly $\alpha$  halo's flux  $F_{\text{halo}}$ ,

$$X_{\text{Ly}\alpha, \text{halo}} = \frac{F_{\text{halo}}}{F_{\text{gal}} + F_{\text{halo}}}, \quad (18)$$

where the integrated Ly $\alpha$  halo flux is given by

$$F_{\text{halo}} = \int \langle \text{SB}_{\alpha}(r_{\perp}) \rangle \frac{2\pi r_{\perp} dr_{\perp}}{D_{\Lambda}^2(z)(1+z)^2}, \quad (19)$$

and the Ly  $\alpha$  line flux of the central galaxy is

$$F_{\text{gal}} = \left\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \right\rangle \langle \mathcal{T}_\alpha \rangle \frac{\langle L_\alpha^{\text{intr}} \rangle}{4\pi D_L^2(z)}. \quad (20)$$

$D_A(z)$  and  $D_L(z)$  are the angular diameter and luminosity distance. Note that in our model the Ly  $\alpha$  halo flux fraction is independent of  $\langle L_\alpha^{\text{intr}} \rangle$  as the increase of the intrinsic Ly  $\alpha$  luminosity increases both the Ly  $\alpha$  halo surface brightness and the observed Ly  $\alpha$  luminosity of galaxies by the same factor. Furthermore, our halo flux fraction model depends on the ISM Ly  $\alpha$  escape fraction in a functional form  $X_{\text{Ly}\alpha,\text{halo}} = 1/(1 + C \langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle)$  where  $C \propto \langle \mathcal{T}_\alpha \rangle / \int \langle \text{SB}_\alpha(r_\perp) \rangle 2\pi r_\perp dr_\perp$  is the CGM/IGM-dependent number. The  $\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle$  parameter only enters in the observed galaxy flux because the H I gas in the dust-free ISM only scatters the Ly  $\alpha$  photons out of an observer's line of sight (but does not permanently destroy the photons), thus the scattered Ly  $\alpha$  photons eventually find the way into the CGM and power Ly  $\alpha$  haloes.

Overall, the above perturbative approach self-consistently models both Ly  $\alpha$  in emission and absorption. The CGM/IGM around galaxies are traced by the cross-correlation between Ly  $\alpha$  forests and galaxies. Thus, armed with this cosmological Ly  $\alpha$  transfer calculation, we can study how the CGM and IGM affect the Ly  $\alpha$  emission-line profiles and Ly  $\alpha$  haloes as a function of the observationally calibrated gaseous environments.

#### 4 PHYSICAL MODEL OF THE GASEOUS ENVIRONMENTS AROUND GALAXIES

Finally, to realize the constrained RT technique we need to specify a parametrized model of the CGM and IGM around galaxies, which will be fitted to Ly  $\alpha$  absorption observations. In this section, we describe a phenomenological model of the CGM and IGM. More specifically, our perturbative approach to Ly  $\alpha$  RT requires models of (i) the H I CDDF,  $\frac{\partial^2 \mathcal{N}}{\partial N_{\text{H I}} \partial z}$ , (ii) the real-space galaxy-absorber correlation function,  $\xi(r)$ , and (iii) the average velocity field,  $\langle v_r(r) \rangle$ , and velocity dispersion,  $\sigma_v(r)$ , between galaxies and absorbers. These models are independent to the perturbative expansion of the cosmological Ly  $\alpha$  transfer introduced in Section 3. Any physical model can be used. To illustrate the joint Ly  $\alpha$  emission-absorption analysis, we adopt a simple model in this paper.

The structure of the CGM/IGM around galaxies is determined by both the structure of the intergalactic filaments in the large-scale structure and the hydrodynamical or radiative interaction of the gas with galaxies. Thus, we divide the modelling of the real-space clustering of the CGM/IGM around galaxies and its kinematics into two contributions: (1) from cosmological structure formation and (2) from the photoionization feedback and galactic winds from galaxies.

##### 4.1 The spatial distribution of gas around galaxies

In our phenomenological model, the real-space correlation function between H I gas and galaxies is divided into the two contributions: (1) from cosmological structure formation (Section 4.1.1) and (2) from the photoionization feedback (Section 4.1.2). The resulting correlation function is expressed as,

$$1 + \xi(r) = C^{\text{phot}}(r) [1 + \xi^{\text{grav}}(r)], \quad (21)$$

where  $\xi^{\text{grav}}(r)$  is the galaxy-absorber correlation function resulting from the gravitational clustering through the structure formation, whereas  $C^{\text{phot}}(r)$  is the correction factor for the clustering due to the photoionization feedback by galaxies.

##### 4.1.1 Gravitational clustering

The gravitational clustering of gas around galaxies is modelled using the power-law form with a slope  $\gamma$  and a correlation length  $r_c(z)$ ,

$$\xi^{\text{grav}}(r, z) = \left[ \frac{r}{r_c(z)} \right]^{-\gamma}. \quad (22)$$

Note that we have adopted the self-similar clustering ansatz, where the time-dependence of the gas-galaxy clustering only enters as a rescaling of the correlation length as a function of redshift,  $r_c(z)$ , while the functional shape is independent of redshift. Although the assumption of the self-similar clustering is only a crude approximation, it permits us to analytically capture the features of the full non-linear growth of the gravitational clustering (Bertschinger 1985; Smith et al. 2003).

##### 4.1.2 Photoionization and UV background

The photoionization of the surrounding gas by the LyC photons ( $\lambda < 912 \text{ \AA}$ ) from the galaxies lowers the number density of the absorbers around galaxies. As shown in Appendix B, the impact of local ionizing sources introduces the correction factor,

$$C^{\text{phot}}(r) = \left[ 1 + \left( \frac{r}{r_{\text{ph}}} \right)^{-2} \right]^{-\beta_{\text{eff}}+1}, \quad (23)$$

where  $\beta_{\text{eff}}$  is the effective slope of the CDDF when fit by a single power law. The equality radius  $r_{\text{ph}}$  is the radius at which the local contribution,  $\Gamma_{\text{local}}(r)$ , to photoionization rate is equal to the average photoionization rate of the UV background.  $\Gamma_{\text{bkg}}$  is the average UV background, for which we assume the value of  $\Gamma_{\text{bkg}} = 1.0 \times 10^{-12} \text{ s}^{-1}$  consistent with the observed value of  $0.86^{+0.30}_{-0.22} \times 10^{-12} \text{ s}^{-1}$  at  $z = 3$  (Becker & Bolton 2013). We use the CDDF function fit by Haardt & Madau (2012) at  $z = 3$  and the effective slope of  $\beta_{\text{eff}} = 1.5$ . Then, the equality radius is given by

$$\begin{aligned} r_{\text{ph}} &= \sqrt{\frac{(1+z)^2}{4\pi\Gamma_{\text{bkg}}} \int_{\nu_{912}}^{\infty} \sigma_{\text{HI}}(\nu) \frac{L_\nu(\nu)}{h\nu} d\nu}, \\ &\approx 133 h^{-1} \text{ ckc} \left( \frac{1+z}{4} \right) \left( \frac{\langle f_{\text{esc}}^{\text{LyC}} \rangle}{0.02} \right)^{\frac{1}{2}} \left( \frac{\langle \text{SFR} \rangle}{34 M_\odot \text{ yr}^{-1}} \right)^{\frac{1}{2}}, \end{aligned} \quad (24)$$

where  $\sigma_{\text{HI}}(\nu) = \sigma_{912}(\nu/\nu_{912})^{-3}$ ,  $\sigma_{912} = 6.304 \times 10^{-18} \text{ cm}^2$ , and  $\nu_{912}$  is the frequency at the Lyman limit. We assumed the spectral energy distribution of  $L_\nu(\nu) = h\alpha \dot{N}_{\text{ion}}(\nu/\nu_{912})^{-\alpha}$  with a power-law EUV slope of  $\alpha = 3$ , and used equation (7) for the LyC photon production rate,  $\dot{N}_{\text{ion}}$ . We use  $r_{\text{ph}} = 133 h^{-1} \text{ ckc}$  (i.e.  $\langle f_{\text{esc}}^{\text{LyC}} \rangle = 0.02$  and  $\langle \text{SFR} \rangle = 34 M_\odot \text{ yr}^{-1}$ ) for the fiducial model.

In addition to the local ionizing sources, the large-scale UV background fluctuation can also modulate the number density of absorbers around galaxies. Because the length-scale of the UV background fluctuations is of order of the mean-free path of ionizing photons,  $\lambda_{912} \approx 346[(1+z)/4]^{-4.4} h^{-1} \text{ cMpc}$  (Worseck et al. 2014) and is much larger than the length-scale of the CGM and IGM around galaxies, we model the impact of the large-scale UV background fluctuations by rescaling the CDDF. Following Hui, Gnedin & Zhang (1997) (see Appendix B), the CDDF responds as a function of the photoionization rate of the UV background,

$$\frac{\partial^2 \mathcal{N}}{\partial N_{\text{H I}} \partial z} = \frac{\partial^2 \mathcal{N}}{\partial N_{\text{H I}} \partial z} \bigg|_{\Gamma_{\text{bkg}}(z)} \left[ \frac{\Gamma_{\text{bkg}}}{\bar{\Gamma}_{\text{bkg}}(z)} \right]^{-\beta_{\text{eff}}+1}. \quad (25)$$



$\bar{\Gamma}_{\text{bkg}}(z)$  is the average UV background of the entire Universe, for which we assume the value of  $\bar{\Gamma}_{\text{bkg}}(z=3) = 1.0 \times 10^{-12} \text{ s}^{-1}$ .

#### 4.2 The kinematics of absorbers around galaxies

The cosmological inflow of the gas onto galaxies is predominantly controlled by the gravitational interaction through the large-scale structure formation. We can then write a Boltzmann equation for the pairs of galaxies and absorbers (Lagrangian gas parcels). By applying the method of BBGKY hierarchy (Davis & Peebles 1977), without loss of generality, the resulting first moment of the BBGKY hierarchy describes the conservation law of the galaxy–absorber pairs. This implies that the cosmological inflow is described by the average velocity field of absorbers around galaxies as

$$\langle v_r(r) \rangle = \frac{H(z)}{1 + \xi^{\text{grav}}(r, z)} \frac{\partial}{\partial z} \int_0^r \xi^{\text{grav}}(r', z) \left(\frac{r'}{r}\right)^2 dr'. \quad (26)$$

While the BBGKY hierarchy is exact, in order to provide an analytic function for the average velocity field, we further assume the self-similar clustering ansatz (Section 4.1). Solving equation (26), we find

$$\langle v_r(r) \rangle = -v_{\text{inflow}} \frac{2(r/r_c)}{1 + (r/r_c)^\gamma}. \quad (27)$$

We introduced a free parameter  $v_{\text{inflow}}$  to parametrize the maximum average inflow velocity of gas onto galaxies.<sup>5</sup> For the velocity dispersion of gas around galaxies, we assume the constant velocity dispersion  $\sigma_v(r) = \text{const}$  for simplicity.

The galactic outflow also affect the gas kinematics around galaxies, especially at inner radii close to the central galaxies. We have tested our model with outflow by launching the wind at a constant velocity at  $\sim 200 \text{ km s}^{-1}$ , which is consistent with the value used for Ly $\alpha$  ISM model of galaxies (Section 3.2.1) and the median outflow velocity measured from the interstellar metal absorption lines (Steidel et al. 2010). The region of mechanical influence is out to radius  $\approx 200 \text{ pkpc}$  which is the maximum radius that gas can be inertially transported during 1 Gyr. The result was only marginally affected; thus, we avoid introducing an extra parameter for outflow in the CGM and IGM in this paper.

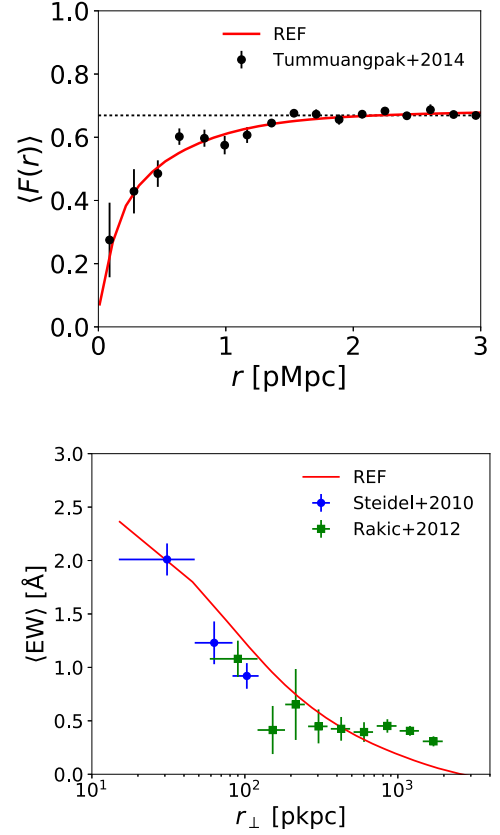
### 5 JOINT ANALYSIS OF GALAXY-LY $\alpha$ ABSORPTION CLUSTERING WITH LY $\alpha$ EMISSION

#### 5.1 Ly $\alpha$ absorption around galaxies

##### 5.1.1 Parameter calibration and observational test

We calibrate our model parameters of the CGM and IGM around galaxies introduced in Section 4 using the observational data from the spectroscopic galaxy surveys in the QSO fields: VLT LBG Redshift Survey (VLRS) (Crighton et al. 2011; Tummuangpak et al. 2014; Bielby et al. 2017), Adelberger et al. (2003, 2005), and Keck Baryonic Structure Survey (Steidel et al. 2010; Rakic et al. 2012;

<sup>5</sup>In the self-similar clustering model, the solution to the BBGKY hierarchy follows  $v_{\text{inflow}} = \frac{1}{2} \frac{\gamma}{3-\gamma} \frac{d \ln r_c}{d \ln a} \frac{H r_c}{1+z}$ , which depends on the time dependence of the correlation length,  $\frac{d \ln r_c}{d \ln a}$ . We expect the value of  $d \ln r_c / d \ln a$  is order unity  $\sim \mathcal{O}(1)$ . For example, in the linear perturbation theory, the correlation length is linearly proportional to the scale length  $d \ln r_c / d \ln a = 1$ . Therefore, the inflow velocity parameter is  $v_{\text{inflow}} \approx 63 \text{ km s}^{-1} (r_c / 1 h^{-1} \text{ cMpc})$  at  $z=3$  and  $\gamma=1.85$ .



**Figure 2.** Comparison of the REF model with data. Top panel: the spherically averaged transmitted Ly $\alpha$  transmitted flux of the REF model (red solid) as a function of radius. The black data points are the measurement from VLRS survey (Tummuangpak et al. 2014). The dotted line is the mean Ly $\alpha$  transmitted flux of the IGM (Becker et al. 2013). Bottom panel: the excess EW of Ly $\alpha$  absorption as a function of impact parameter  $r_{\perp}$  for the REF model (red solid). The blue and green data points are the measurements from galaxy–galaxy pairs (Steidel et al. 2010) and galaxy–QSO pairs (Rakic et al. 2012) in Keck Baryonic Structure Survey.

Turner et al. 2014). We choose the best-fitting correlation length  $r_c$  and slope  $\gamma$ , the inflow velocity parameter  $v_{\text{inflow}}$ , and the velocity dispersion parameter  $\sigma_v$ , by comparing the model prediction of the 2D galaxy–Ly $\alpha$  forest cross-correlation function with observed one. We also use the EW of Ly $\alpha$  absorption around galaxies by Steidel et al. (2010) and Rakic et al. (2012) to complement the calibration procedure. We then find the best-fitting parameters by fitting the model by eye.

The CDDF,  $\frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z}$ , and photoionization rate of the UV background,  $\Gamma_{\text{bkg}}$ , are pre-determined based on the observation and analysis of QSO absorption spectra (see Section 4.1).

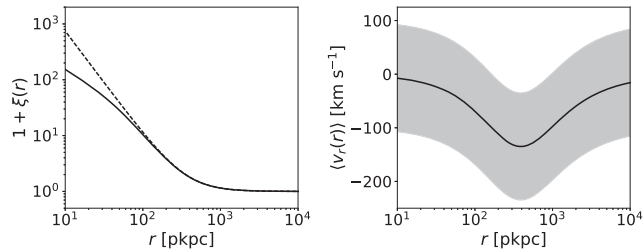
To find the best-fitting parameters of  $r_c$  and  $\gamma$ , we first compare the model with the spherically averaged Ly $\alpha$  transmitted flux and the EW of Ly $\alpha$  absorption. Fig. 2 shows the result of the calibration. The resulting parameters are listed in Table 1 and we refer to this set of parameters as REF model in the rest of the paper.

Fig. 3 (left-hand panel) plots the best-fitting real-space correlation function between absorbers and galaxies (solid curve), which include both contributions from gravitational clustering and photoionization feedback. As a reference, the dashed curve shows the correlation function without photoionization feedback. The clustering of the gas around galaxies increases the abundance of the



**Table 1.** The model parameters of star-forming galaxies and the CGM/IGM.

ISM parameters	
SFR	$\langle \text{SFR} \rangle = 34 M_{\odot} \text{ yr}^{-1}$
Average ISM Ly $\alpha$ escape fraction	$\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle = 0.20$
Average LyC escape fraction	$\langle f_{\text{esc}}^{\text{LyC}} \rangle = 0.02$
ISM Ly $\alpha$ line profile	Gronke & Dijkstra (2016) model
CGM/IGM parameters	
Correlation length	$r_c = 1.0 h^{-1} \text{ cMpc}$
Power-law slope	$\gamma = 1.85$
Photoionization rate of UV background	$\Gamma_{\text{bkg}} = 1 \times 10^{-12} \text{ s}^{-1}$
Inflow velocity	$v_{\text{inflow}} = 135 \text{ km s}^{-1}$
Velocity dispersion	$\sigma_v = 200 \text{ km s}^{-1}$



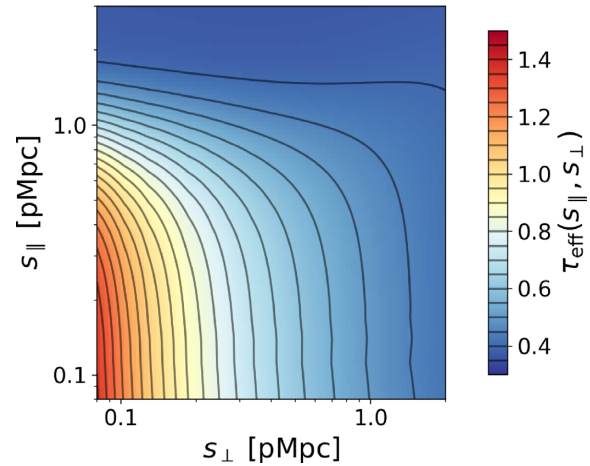
**Figure 3.** The reference model of the CGM and IGM around galaxies. Left-hand panel: the real-space correlation function of H I gas around galaxies of the REF model with the photoionization feedback (solid). The dotted line shows the case for the gravitational clustering without photoionization feedback. Right-hand panel: the average peculiar velocity field of gas around galaxies. The solid line is the cosmological inflow based on self-similar clustering ansatz. The width of the grey shaded region indicates  $1\sigma$  scatter due to the velocity dispersion in the Gaussian streaming model.

absorbers, which extends out to  $\sim 1$  pMpc. Compared with the purely gravitational contribution, the photoionization feedback by the central galaxies mildly lowers the abundance of the neutral gas within  $\lesssim 100$  pkpc.

The gas kinematics is poorly constrained by the available Ly  $\alpha$  absorption data. While VLRS reports the measurement of the 2D galaxy-Ly  $\alpha$  forest cross-correlation at  $z = 3$  (Tummuangpak et al. 2014; Bielby et al. 2017), the tabulated data to be directly compared with the model was not available at the time of writing. Rakic et al. (2012) and Turner et al. (2014) present the tabulated data, but in terms of the median optical depth. As we cannot directly compare our model with these measurements, we choose an inflow velocity parameter,  $v_{\text{inflow}}$ , by comparing with cosmological hydrodynamic simulations of Meiksin, Bolton & Tittley (2014) and Meiksin et al. (2015) that are tested against the observed Ly  $\alpha$  absorption signal around galaxies (Rakic et al. 2012). Meiksin et al. (2014) report the mean radial velocity profile of gas around haloes with total mass of  $4.5 \times 10^{11} M_{\odot}$  at  $z = 3$ . The simulations show the gas inflow of velocity  $\sim 100 - 150 \text{ km s}^{-1}$  around  $\sim 1 - 10 \text{ cMpc}$ . We therefore choose the inflow velocity parameter of  $v_{\text{inflow}} = 135 \text{ km s}^{-1}$  for our REF model.<sup>6</sup>

Furthermore, we choose the constant velocity dispersion parameter of  $\sigma_v = 200 \text{ km s}^{-1}$ . This is based on the reported value of

<sup>6</sup>Near the completing of our work, Turner et al. (2017) independently reported a comparison of the observed 2D median optical depth map with the EAGLE simulations, and find a similar inflow velocity (velocity dispersion) to our phenomenological model. This assures the fidelity of our model and calibration.



**Figure 4.** 2D Ly  $\alpha$  effective optical depth map around galaxies at  $z = 3$  as a function of the line-of-sight,  $s_{\parallel}$ , and the transverse,  $s_{\perp}$ , separations between the Ly  $\alpha$  absorption pixels and galaxies in the REF model.

$240 \pm 60 \text{ km s}^{-1}$  by Bielby et al. (2017) derived by their model fitting procedure to the 2D galaxy-Ly  $\alpha$  forest cross-correlation. Once the tabulated data of the 2D galaxy-Ly  $\alpha$  forest cross-correlation is available, we can readily calibrate our model of the gas kinematics in a more consistent manner in future.

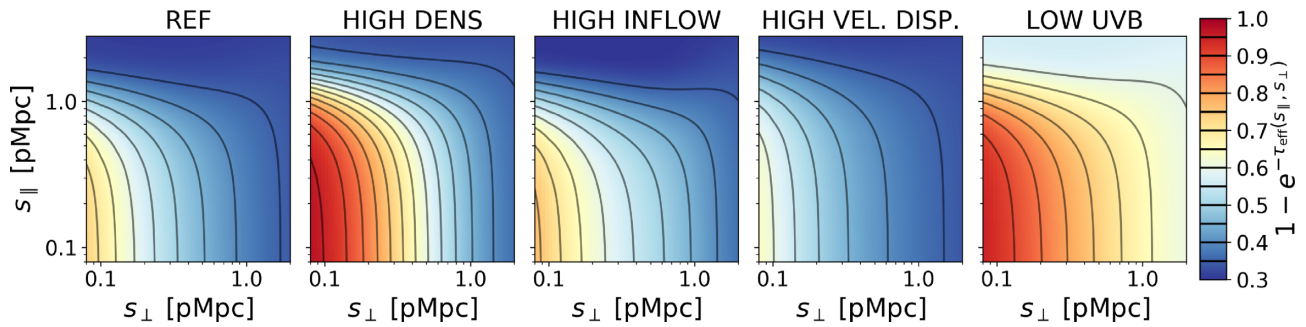
Fig. 3 (right-hand panel) shows the average peculiar velocity field of gas around galaxies (solid curve) in our REF model. The grey shaded region indicates the velocity dispersion of the gas. The gas infalls from large to small scales by cosmological inflow as a result of structure formation. The mean peculiar gas velocity then approaches zero at inner radii inside the turnaround radius of the haloes. This trend is in agreement with the simulations of van de Voort & Schaye (2012) and Meiksin et al. (2014), where such deceleration is caused by the shocks across the multistreaming gas and the virialization of the haloes.

After the calibration, our model of the gas distribution and kinematics of the CGM and IGM around galaxies are constrained to be consistent with the galaxy-Ly  $\alpha$  forest clustering data. Thus, the Ly  $\alpha$  RT calculation using this calibrated model allows us to self-consistently study the interaction between Ly  $\alpha$  emission-line profiles and haloes of galaxies and the large-scale gaseous environment.

### 5.1.2 2D redshift-space galaxy-Ly $\alpha$ forest clustering: the imprints of the CGM and IGM

Before proceeding to the Ly  $\alpha$  emission properties, we present the 2D redshift-space galaxy-Ly  $\alpha$  forest clustering, which contains the full statistical information of the clustering and kinematics of the CGM and IGM around galaxies. Fig. 4 shows the 2D effective optical depth map  $\tau_{\text{eff}}(s_{\parallel}, s_{\perp}) = -\ln \langle F(s_{\parallel}, s_{\perp}) \rangle$  in the REF model calibrated in the previous section. The model clearly shows the excess of Ly  $\alpha$  absorption near the central galaxies and the redshift-space anisotropy in the 2D effective optical depth map. The feature of the redshift-space anisotropy in the model strikingly resembles the one observed by Turner et al. (2014).

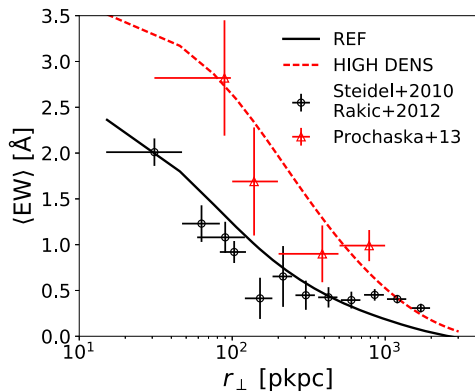
With Fig. 5, we show how the gas distribution and kinematics of the CGM and IGM around galaxies are traced by the 2D redshift-space galaxy-Ly  $\alpha$  forest clustering. The variation of the 2D effective optical depth maps for the four different gaseous environments are shown in Fig. 5. Each model is perturbed relative to the REF model



**Figure 5.** 2D map of Ly $\alpha$  transmitted flux around galaxies. All the outermost contour is  $1 - e^{-\tau_{\text{eff}}} = 0.35$  and increases by 0.05 inwards, except for LOW UVB model for which the outermost contour is 0.60. The model parameters for REF model is shown in Table 1. HIGH DENS model has  $r_c = 2 h^{-1}$  cMpc, HIGH INFLOW model has  $v_{\text{inflow}} = 300 \text{ km s}^{-1}$ , HIGH VEL. DISP. model has  $\sigma_v = 300 \text{ km s}^{-1}$ , and LOW UVB model has  $\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$ . All other parameters are the same as REF model.

**Table 2.** A grid of the CGM/IGM models. The parameter only differs from the REF model is shown.

HIGH DENS	$r_c = 2.0 h^{-1} \text{ cMpc}$
HIGH INFLOW	$v_{\text{inflow}} = 300 \text{ km s}^{-1}$
HIGH VEL. DISP.	$\sigma_v = 300 \text{ km s}^{-1}$
LOW UVB	$\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$



**Figure 6.** Same as the bottom panel of Fig. 2. The figure shows that the HIGH DENS (red dashed) model represents a high-density environment similar to the Prochaska et al. (2013) observation (red triangles).

with respect to gas density, HIGH DENS ( $r_c = 2 h^{-1}$  cMpc), inflow velocity, HIGH INFLOW ( $v_{\text{inflow}} = 300 \text{ km s}^{-1}$ ), and velocity dispersion, HIGH VEL. DISP. ( $\sigma_v = 300 \text{ km s}^{-1}$ ), and UV background, LOW UVB ( $\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$ ) (tabulated in Table 2). All the other parameters are the same as the REF model.

The H I gas density affects the amplitude of the 2D effective optical depth map. Our HIGH DENS model in Fig. 5 shows that increasing the correlation length of gas around galaxies increases the effective optical depth near the centre. Such high-density environment resembles the region around QSOs. This is illustrated in Fig. 6 where we compare the Ly $\alpha$  absorption EW in our HIGH DENS model with the observation around QSOs (Prochaska et al. 2013). The photoionization similarly affects the amplitude (see LOW UVB model). While a higher LyC leakage from galaxies preferentially lowers the Ly $\alpha$  absorption closer to galaxies, a lower photoionization rate of the UV background uniformly increases the opacity of gas around galaxies, making the CGM and IGM more neutral.

On the other hand, the gas kinematics (inflow and velocity dispersion: HIGH INFLOW and HIGH VEL. DISP. models) impact

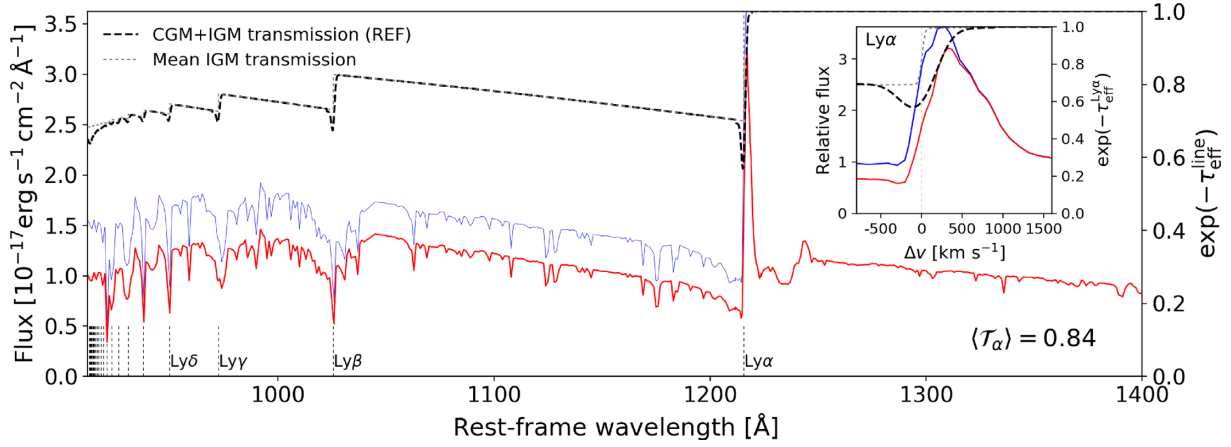
the redshift-space anisotropy. On  $\sim 1$  pMpc scale, the clumpy neutral gas inflow onto galaxies causes the squashing of the effective optical depth map along the line-of-sight direction. Such large-scale inflow of the gas on to galaxies has been detected in observations (Rakic et al. 2012; Bielby et al. 2017) in a consistent matter with cosmological simulations (Turner et al. 2017). A higher velocity dispersion introduces a more line-of-sight elongation in the redshift-space anisotropy. Note that including outflow in the model will also introduce a similar elongation. However, in detail, the quantitative elongation signature by the increasing random gas motion differs from the coherent change in the velocity field by outflow. They should be distinguishable in principle.

## 5.2 Emergent Ly $\alpha$ emission-line profile

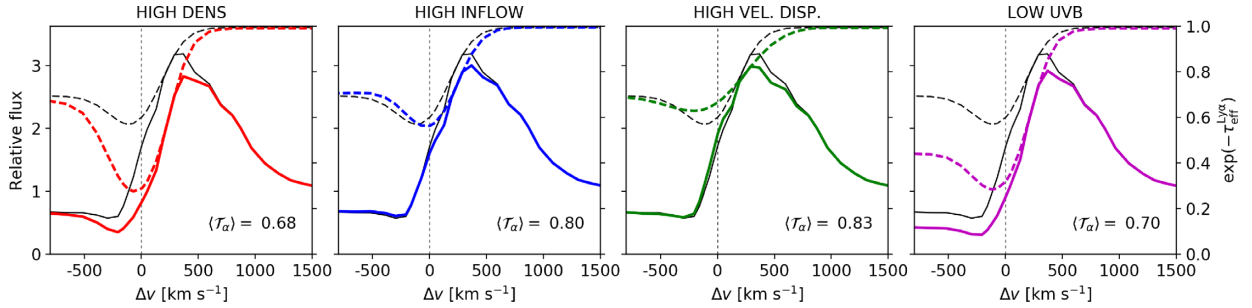
### 5.2.1 The impact of the CGM and IGM predicted from the galaxy-Ly $\alpha$ absorption clustering data

Once the model of the CGM and IGM is constrained by the galaxy-Ly $\alpha$  forest clustering measurements, the constrained Ly $\alpha$  RT approach self-consistently predicts the impact of the CGM/IGM on the emergent Ly $\alpha$  line profile of galaxies. Fig. 7 shows the emergent average spectrum of star-forming galaxies and the Ly $\alpha$  line profile after propagating through the CGM and IGM in the REF model (red solid curve) [the intrinsic model galaxy spectrum (blue solid curve) is shown for a comparison]. The impact of the CGM/IGM is encapsulated in the CGM/IGM transmission curve (black dashed curve with right ordinate), which is a self-consistent estimate of  $e^{-\tau_{\text{eff}}^{\text{Ly}\alpha}}$  (and  $e^{-\tau_{\text{eff}}^{\text{line}}}$ ) (Section 3.2.2) from the galaxy-Ly $\alpha$  forest clustering constraint without any adjustable free parameter.

In the observationally constrained CGM/IGM, only  $\langle \mathcal{T}_\alpha \rangle = 0.84$  of the intrinsic Ly $\alpha$  flux is transmitted, while nearly 100 per cent of photons can propagate toward observers in the absence of clustering (Mean IGM, grey dotted curve). This means that the impact of the CGM/IGM on the Ly $\alpha$  line flux is small, but non-negligible even at  $z \sim 2-3$ . While the precise value of the transmission of course depends on the ISM Ly $\alpha$  line profile as  $\langle \mathcal{T}_\alpha \rangle = \int e^{-\tau_{\text{eff}}^{\text{Ly}\alpha}} \langle \Phi_\alpha^{\text{ISM}} \rangle dv_e$ , this conclusion still holds. We will return to this point in Section 5.3. The inset of Fig. 7 shows in detail the CGM/IGM transmission curve near the Ly $\alpha$  line. The clustering of neutral gas around galaxies scatters more Ly $\alpha$  photons out of our line of sight, causing the ‘attenuation dip’ near the Ly $\alpha$  resonance line centre (Laursen et al. 2011). The shape of this attenuation dip is determined by the H I gas distribution and kinematics of the CGM. We will examine this in detail in the following section.



**Figure 7.** The composite spectrum (red solid, the flux in the left y-axis) of galaxies with the emergent Ly  $\alpha$  line profile and the effective CGM/IGM transmission  $e^{-\tau_{\text{eff}}^{\text{line}}}$  (black dashed, the value is in the right y-axis) in the REF model. The blue line shows the spectrum without the effect of the CGM and IGM, using the BPASS population synthesis model (Binary Population Spectral Synthesis code, Eldridge et al. 2017) with constant SFR, 100 Myr age, and metallicity  $Z = 0.001$  as an example. For a comparison, the effective IGM transmission without the CGM effect (Meiksin 2006) is shown as the grey dotted line (mean IGM model). The value of Ly  $\alpha$  transmission ( $\langle \mathcal{T}_\alpha \rangle$ ) is quoted at the bottom right corner. The inset shows the zoom-in plot of the Ly  $\alpha$  line profile.



**Figure 8.** Comparison of the average emergent Ly  $\alpha$  line profiles in HIGH DENS (red solid), HIGH INFLOW (blue solid), HIGH VEL. DISP. (green solid), and LOW UVB (purple solid) models with REF model (black solid). The value of the relative flux in each line profile is shown in the left y-axis. The CGM/IGM transmission curve for each model is indicated as dashed lines with a value shown in the right y-axis. The value of Ly  $\alpha$  transmission in each model is shown at the bottom right corner. HIGH DENS model has  $r_c = 2 h^{-1} \text{ cMpc}$ , HIGH INFLOW model has  $v_{\text{inflow}} = 300 \text{ km s}^{-1}$ , HIGH VEL. DISP. model has  $\sigma_v = 300 \text{ km s}^{-1}$ , and LOW UVB model has  $\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$ . All other parameters are the same as REF model.

Furthermore, the clustering of gas in the CGM imprints the excess attenuation near the higher order Lyman series in the average galaxy spectrum, adding a series of attenuation dips by the CGM (Rudie et al. 2013) on top of the characteristic sawtooth-shaped IGM transmission curve (Madau 1995; Meiksin 2006; Inoue et al. 2014). The shape of the attenuation dips near the higher order Lyman series is well correlated with that of Ly  $\alpha$ , which only differs by the difference in the oscillator strengths of the atomic transitions. Note that the intrinsic absorption at stellar atmosphere of massive stars (in the stellar population synthesis model) also causes the attenuation at the higher order Lyman series lines. Thus, the scattering of the higher order Lyman series photons by the CGM causes excess attenuations at the higher order Lyman series over the intrinsic stellar atmosphere and ISM features. The higher order Lyman series features in the average spectrum may be used as a consistency check of the impact of the CGM and IGM on the emergent Ly  $\alpha$  line.

### 5.2.2 Dissecting the impacts of the CGM and IGM: the clustering and kinematics of the gas around galaxies

We now examine the details of how the changes of gas density, and kinematics, and UV background in the CGM and IGM around

galaxies traced by the galaxy-Ly  $\alpha$  forest clustering (Fig. 5) impact the Ly  $\alpha$  line profiles. Fig. 8 shows the emergent Ly  $\alpha$  line profiles after propagating through the high gas density (HIGH DENS), high inflow velocity (HIGH INFLOW), high velocity dispersion (HIGH VEL. DISP.), and low UV background (LOW UVB) model environments.

A stronger clustering of gas around galaxies enhances the attenuation of Ly  $\alpha$  just blueward of the line centre, which is responsible for creating a dip of the attenuation in the transmission curve. Ly  $\alpha$  line transfer is sensitive to total velocity field of the gas (i.e. Hubble flow + peculiar velocity). Including the Hubble flow, most of the gas is in fact experiencing net outflow. As the excess distribution of gas by the clustering extends out to  $\sim 1 \text{ pMpc}$  around galaxies where the Hubble flow is  $\sim 300 \text{ km s}^{-1}$  at  $z \sim 3$ , the dip ranges approximately over  $\Delta v \approx 0\text{--}300 \text{ km s}^{-1}$ . The attenuation dip increases with increasing gas density around galaxies. In a high-density environment, the Ly  $\alpha$  flux is suppressed even more. Only  $\langle \mathcal{T}_\alpha \rangle = 0.68$  of the Ly  $\alpha$  flux is transmitted through the CGM and IGM in the high-density environment, while  $\approx 0.84$  of Ly  $\alpha$  flux can be transmitted through the REF model of the CGM and IGM around galaxies.

The (wavelength-dependent) shape of the CGM/IGM transmission curve near the Ly  $\alpha$  line centre is affected by the H I gas kine-



matics around galaxies. The inflowing gas causes the attenuation in the redward ( $\Delta v > 0$ ) of the line centre whereas the outflowing gas attenuates the blueward ( $\Delta v < 0$ ). As the gas infalls onto galaxies, the velocity field deviates from the Hubble flow, resulting the net inflow of the gas. This inflowing gas can scatter Ly $\alpha$  photons intrinsically emitted at the redward of the line from the central galaxies as the photon is seen blueshifted to the line centre in the rest frame of the inflowing gas. Therefore, as the inflowing gas attenuates the Ly $\alpha$  photons, it produces an excess Ly $\alpha$  attenuation at the redward of the line centre. Thus, a higher gas inflow velocity suppresses more Ly $\alpha$  flux at the redward of the line centre. For example, increasing the inflow velocity parameter from  $135 \text{ km s}^{-1}$  to  $v_{\text{inflow}} = 300 \text{ km s}^{-1}$  (HIGH INFLOW model) increases the attenuation of the redshifted Ly $\alpha$  line to  $\langle \mathcal{T}_\alpha \rangle \approx 0.80$ , making the CGM and IGM more susceptible of scattering the redshifted Ly $\alpha$  photons emitted from galaxies.

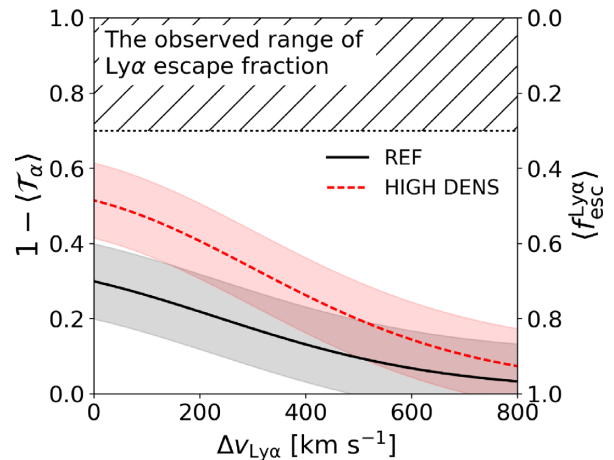
Furthermore, the velocity dispersion of the gas around galaxies impacts the width of the attenuation dip. Because of the velocity dispersion of gas around galaxies, there are both outflowing and infalling gas along with the coherent flow. Such fast moving outliers of the gas broaden the width of the attenuation dip. A higher gas velocity dispersion of the CGM and IGM lowers the attenuation close to the line centre. This is because the chaotic motion of the gas creates the path of escapes in the gas around galaxies, which is otherwise opaque to Ly $\alpha$  photons because of the coherent inflowing gas. On the other hand, at larger radii, where the gas is predominantly outflowing, a higher velocity dispersion means that some gas can be blueshifted into the line centre and scatter the Ly $\alpha$  photons, otherwise the coherent Hubble flow lets the photons to escape. This creates more attenuation at the redward of the line centre than the lower velocity dispersion case.

Finally, a change in the UV background affects the overall Ly $\alpha$  transmission through the CGM and IGM around galaxies. A lower photoionization rate makes gas to be more neutral. This increase in the residual neutrality of the gas increases the abundance of the H I absorbers around galaxies. Because the effective optical depth scales as  $\tau_{\text{eff}}^{\text{Ly}\alpha} \propto \Gamma_{\text{bkg}}^{-1/2}$ , a lower photoionization rate of the UV background attenuates more Ly $\alpha$  photons from galaxies. For example, lowering the photoionization rate from  $\Gamma_{\text{bkg}} = 1 \times 10^{-12}$  to  $\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$  decreases the Ly $\alpha$  transmission from  $\langle \mathcal{T}_\alpha \rangle \approx 0.84$  to 0.70.

### 5.2.3 Relative contribution of absorbers

Different types of absorbers, Ly $\alpha$  forest absorbers with  $N_{\text{H I}} < 10^{17} \text{ cm}^{-2}$  and Lyman-limit systems/damped Ly $\alpha$  absorbers (LSS/DLAs) with  $N_{\text{H I}} > 10^{17} \text{ cm}^{-2}$ , contribute differently to the CGM/IGM transmission curve. At the redward of Ly $\alpha$  line centre, the formation of the red damping wing is mainly driven by the two contributions: (1) infalling Ly $\alpha$  forest absorbers onto galaxies and (2) the LLS/DLAs around galaxies.

As discussed in Section 5.2.2, the infalling absorbers scatter Ly $\alpha$  photons emitted at the redward of line centre. The contribution to the transmission curve extends to the maximum velocity of infalling Ly $\alpha$  forest absorbers. This could be larger than the maximum mean inflow velocity parameter  $v_{\text{inflow}}$  due to the velocity dispersion. As a result of the chaotic gas kinematics (large  $\sigma_v$ ) in the CGM, the CGM/IGM transmission curves extend more smoothly to larger positive  $\Delta v$  than a case of the coherent gas kinematics with a lower velocity dispersion (small  $\sigma_v$ ). In fact, at  $z = 3$ , the contribution of



**Figure 9.** The contribution of the CGM and IGM on the Ly $\alpha$  escape fraction at  $z \sim 2-3$ . The black solid (red dashed) curve shows the Ly $\alpha$  transmission of the CGM and IGM,  $\langle \mathcal{T}_\alpha \rangle$ , in the REF (HIGH DENS) model as a function of Ly $\alpha$  velocity offset. The grey and red shaded regions indicate the typical model uncertainty bracketing the line width between 100 and 400  $\text{km s}^{-1}$ . The hatched region indicates the observed range of Ly $\alpha$  escape fraction derived from the Ly $\alpha$ /H $\alpha$  ratio. The right y-axis shows the value of Ly $\alpha$  escape fraction if the CGM/IGM is only the source of opacity, i.e.  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle = \langle \mathcal{T}_\alpha \rangle$ . The figure illustrates that to explain the observed Ly $\alpha$  escape fraction, galaxies should have a large contribution from the ISM, but there is a non-negligible impact from the CGM and IGM.

infalling Ly $\alpha$  forest absorbers dominates the formation of the red damping wing opacity (see Appendix D).

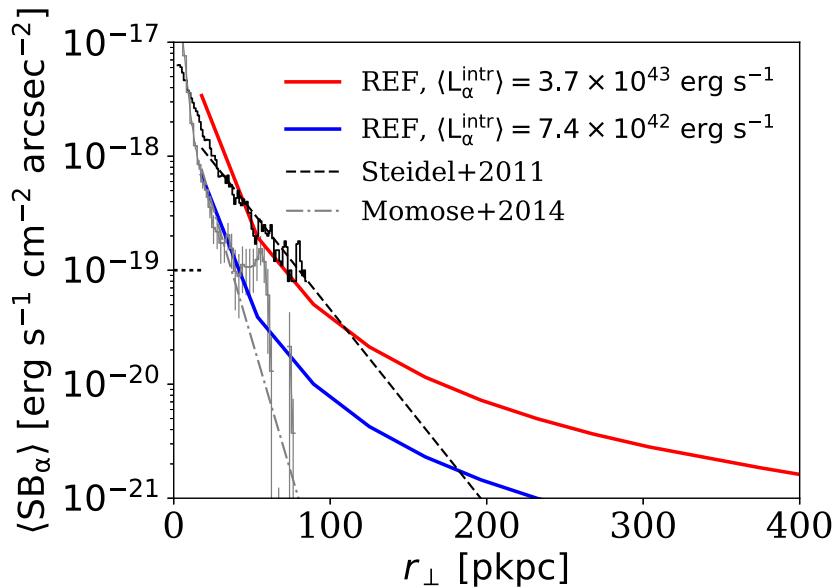
While the Ly $\alpha$  forest absorbers contribute to the red damping wing when the gas is inflowing, the LLS/DLAs around galaxies can contribute the damping wing even when the gas is outflowing. Since the LLS/DLAs show the Lorentz wing absorption due to the strong Ly $\alpha$  absorption, the Lorentz wing scattering by the outflowing LLS/DLAs contributes to the redward of line centre. This effect is subdominant in the damping wing at  $z = 3$  because the number density of LLS/DLAs is low. However, when the photoionization rate of the UV background is lower, the increasing number density of LLS/DLAs causes a non-negligible contribution, and can dominate the formation of the red damping wing opacity. The result is presented in Appendix D.

### 5.3 Ly $\alpha$ escape fraction

We now examine the impact of the CGM and IGM on Ly $\alpha$  escape fraction. Observationally,  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle$  can be estimated from galaxy spectra using the ratio between the observed Ly $\alpha$  line flux and the expected Ly $\alpha$  line flux inferred from the observed H $\alpha$  line flux, which are both produced following recombination with a known ratio. The Ly $\alpha$  flux is typically suppressed relative to this ratio, which reflects either that Ly $\alpha$  photons were efficiently destroyed by dust, or that Ly $\alpha$  photons were scattered into an extended low surface brightness halo. In practice, observations generally cannot distinguish between these two scenarios, while theoretically we expect both physical processes to affect the observationally inferred  $f_{\text{esc}}^{\text{Ly}\alpha}$ .

Fig. 9 shows the contribution of the CGM and IGM to the Ly $\alpha$  escape fraction in our model as a function of Ly $\alpha$  velocity offset from line centre. The black solid (red dashed) curve shows the total Ly $\alpha$  escape fraction only taking into account the contribution from the CGM and IGM, i.e.  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle = \langle \mathcal{T}_\alpha \rangle$ , in the REF model (HIGH





**Figure 10.** Mean surface brightness profile of Ly  $\alpha$  haloes in the REF model with two different intrinsic Ly  $\alpha$  luminosities ( $L_{\alpha}^{\text{intr}} = 3.7 \times 10^{43} \text{ erg s}^{-1}$  (red curve) and  $7.4 \times 10^{42} \text{ erg s}^{-1}$  (blue curve)). The black and grey histograms show the observed surface brightness profiles of Ly  $\alpha$  haloes around LBGs (Steidel et al. 2011) and LAEs (Momose et al. 2014); the dashed and dash-dotted lines are their best-fitting exponential profiles. The horizontal dotted line indicates the approximate surface brightness limit ( $10^{-19} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}$ ) of the observations.

DENS model). In Fig. 9 also compares these curves to the typical observed range (hatched region) (e.g. Hayes et al. 2010; Trainor et al. 2015). This shows that the contribution from the CGM and IGM alone is not sufficient to explain the observed Ly  $\alpha$  escape fraction, *but the CGM has a non-negligible impact*. In addition, the contribution of the CGM to the Ly  $\alpha$  escape fraction is larger at higher CGM densities (HIGH DENS model), which can introduce an environmental impact on the observationally inferred Ly  $\alpha$  escape fraction.

While the importance of the ISM controlling the Ly  $\alpha$  escape is well appreciated in literature, our analysis provides independent support of this conclusion based on the galaxy-Ly  $\alpha$  forest clustering data. However, our results imply that Ly  $\alpha$  RT does not end after Ly  $\alpha$  photons escape from galaxies, even at  $z \sim 2-3$ .

In detail, Ly  $\alpha$  escape from the ISM and Ly  $\alpha$  RT through the CGM/IGM are coupled: scattering in the ISM affects the velocity offset of Ly  $\alpha$  from line centre (Verhamme, Schaerer & Maselli 2006; Gronke & Dijkstra 2016). The observed range of Ly  $\alpha$  velocity offsets,  $\Delta v_{\text{Ly}\alpha} = 0 - 600 \text{ km s}^{-1}$  (Steidel et al. 2010) gives rise to IGM/CGM transmissions in the range  $\langle T_{\alpha} \rangle \approx 0.70$  to  $\approx 0.90$  (REF model). Scattering through galactic outflows can give rise to large intrinsic Ly  $\alpha$  velocity offsets, where the CGM and IGM are more transparent (even during the reionization epoch, see Dijkstra & Wyithe 2010). The environmental dependence of IGM/CGM transmission can have interesting implications for the large-scale clustering of Ly  $\alpha$  emitters (LAEs) and the possible non-gravitational contribution by the Ly  $\alpha$  RT effect through the surrounding gas environments around Ly  $\alpha$  emitting galaxies (see Section 6.2).

## 5.4 Ly $\alpha$ haloes

### 5.4.1 Self-consistent profile of Ly $\alpha$ haloes based on the galaxy-Ly $\alpha$ absorption clustering data

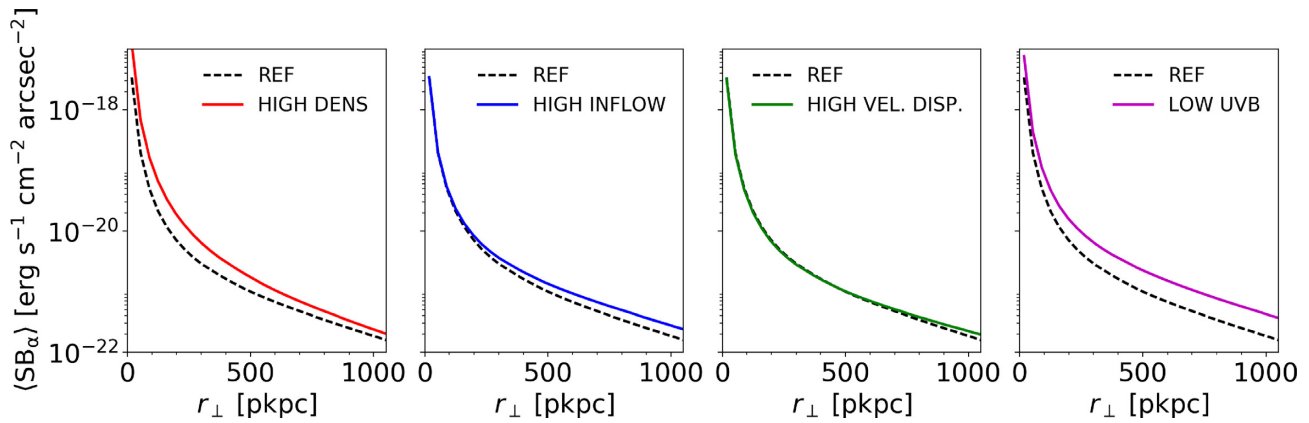
Finally, we present our results for the Ly  $\alpha$  haloes based on the galaxy-Ly  $\alpha$  forest clustering data. Fig. 10 shows the model predic-

tion of the mean surface brightness profile of Ly  $\alpha$  haloes around star-forming galaxies. The red and blue curves show the Ly  $\alpha$  haloes for two different intrinsic Ly  $\alpha$  luminosities of the central galaxies,  $\langle L_{\alpha}^{\text{intr}} \rangle = 3.7 \times 10^{43}$  and  $7.4 \times 10^{42} \text{ erg s}^{-1}$ . The former is our fiducial model with  $\text{SFR} = 34 M_{\odot} \text{ yr}^{-1}$ . The latter shows a case for  $\text{SFR} = 6.8 M_{\odot} \text{ yr}^{-1}$ , illustrating a dependence of Ly  $\alpha$  haloes on the star formation activity of galaxies.

Comparing with the observed mean surface brightness profiles of LBGs (black histogram, Steidel et al. 2011) and LAEs (grey histogram, Momose et al. 2014), the model reproduces the mean surface brightness profile well in the observed range  $r_{\perp} \approx 20 - 80 \text{ pkpc}$ . We emphasize that because the distribution and kinematics of the CGM and IGM are pre-constrained by the galaxy-Ly  $\alpha$  forest clustering data (Section 5.1), we did not adjust any parameter regarding the structure of the gaseous environment. At a fixed intrinsic Ly  $\alpha$  luminosity, the model can be considered as a self-consistent and unique prediction of the mean surface brightness profile of Ly  $\alpha$  haloes for a given formation mechanism (Ly  $\alpha$  scattering of the central sources).

The clustering of neutral hydrogen gas around galaxies is responsible for producing the Ly  $\alpha$  haloes. Ly  $\alpha$  scattering in the mean IGM will underestimate the surface brightness of Ly  $\alpha$  haloes. As the amount of neutral gas increases at inner radii, the intrinsic Ly  $\alpha$  emission from the central galaxies are more likely to be scattered back onto the line of sight. While this decreases the visibility of the Ly  $\alpha$  line in the galaxy spectra, these attenuated Ly  $\alpha$  photons are not permanently lost. Some of the photons are scattered back toward observers and seen as the diffuse Ly  $\alpha$  haloes around the galaxies. Furthermore, since we assume the powering by central star-forming galaxies, the normalization of the surface brightness profile scales with the intrinsic Ly  $\alpha$  luminosity,  $\langle SB_{\alpha} \rangle \propto \langle L_{\alpha}^{\text{intr}} \rangle$ .<sup>7</sup>

<sup>7</sup>Here, the intrinsic Ly  $\alpha$  luminosity refers to the total amount of Ly  $\alpha$  photons leaked out to the CGM and IGM from the ISM of galaxies. We note that the surprisingly good match to the observed Ly  $\alpha$  haloes requires all



**Figure 11.** Comparison of the mean surface brightness profiles of Ly $\alpha$  haloes in the REF model (black dashed curve) with HIGH DENS, HIGH INFLOW, HIGH VEL. DISP., and LOW UVB models (solid coloured curves from left to right). The model parameters for REF model is shown in Table 1. HIGH DENS model has  $r_c = 2 h^{-1} \text{ cMpc}$ , HIGH INFLOW model has  $v_{\text{inflow}} = 300 \text{ km s}^{-1}$ , HIGH VEL. DISP. model has  $\sigma_v = 300 \text{ km s}^{-1}$ , and LOW UVB model has  $\Gamma_{\text{bkg}} = 2 \times 10^{-13} \text{ s}^{-1}$ . All other parameters are the same as REF model.

The scaling with SFR can explain the difference between Ly $\alpha$  haloes around LBGs and LAEs. The surface brightness profile of Ly $\alpha$  haloes produced by scatterings of Ly $\alpha$  photons can be fitted by a power law,

$$(\text{SB}_\alpha(r_\perp)) / [\text{erg s}^{-1} \text{ cm}^{-2} \text{ arcsec}^{-2}] \approx 2.1 \times 10^{-18} \left( \frac{\langle L_\alpha^{\text{intr}} \rangle}{3.7 \times 10^{43} \text{ erg s}^{-1}} \right) \left( \frac{r_\perp}{20 \text{ pkpc}} \right)^{-2.4}. \quad (28)$$

This fit is accurate to 10 per cent relative to the direct result from the model at  $20 \text{ pkpc} < r_\perp < 1000 \text{ pkpc}$ .

The surface brightness profile also depends on the intrinsic Ly $\alpha$  velocity offset. When the intrinsic Ly $\alpha$  line profile has a smaller Ly $\alpha$  velocity offset, close to the line centre, the Ly $\alpha$  photons have more probability to be scattered by the surrounding CGM and IGM, increasing the surface brightness. However, this is a secondary effect compared to a dominating change caused by the intrinsic Ly $\alpha$  luminosity.

Over the radius 20–1000 pkpc, the mean surface brightness profile is well described by the power law  $\propto r^{-\alpha}$  with the slope of  $\alpha \approx -2.4$ . The power law tail extends out to  $\sim 1 \text{ pMpc}$  in the model prediction. This is in contrast with the conventional exponential profile  $\propto \exp(-r/r_1)$  with a scale length  $r_1$ . Although within the observed range below 80 pkpc both the power-law and exponential profiles describe the mean surface brightness profile of Ly $\alpha$  haloes, in the outskirts of the CGM, the exponential profile substantially underestimates the Ly $\alpha$  surface brightness. Requiring the self-consistency with the galaxy-Ly $\alpha$  forest clustering data, the formation of Ly $\alpha$  haloes powered by the scattering of Ly $\alpha$  photons from central star-forming galaxies predicts that there must be large extent of diffuse Ly $\alpha$  emission from the CGM and IGM around galaxies. If this picture that Ly $\alpha$  haloes are powered by scattering

Ly $\alpha$  photons produced within the ISM eventually leak out to the CGM. The effect of the absorption by dust enters as a lower value of the intrinsic Ly $\alpha$  luminosity. Therefore, we are not concluding that the Ly $\alpha$  scattering is the origin of Ly $\alpha$  haloes. Instead, we conclude that our joint Ly $\alpha$  emission – absorption modelling is consistent with the observation as the predicted contribution from Ly $\alpha$  scatterings is within the observed Ly $\alpha$  surface brightness. Nevertheless, despite the caveat, the remarkable match to the observation is worth noting.

is true, deeper observation should find an extended diffuse tail of Ly $\alpha$  haloes.

We argue that the prediction of the extended power-law tail of Ly $\alpha$  emission in the diffuse haloes is a robust result *required* from the galaxy-Ly $\alpha$  absorption clustering data; the profile is difficult to change by changing the ISM and star-forming properties of the galaxies such as the ISM Ly $\alpha$  escape fraction and SFR, which only enters as a rescaling of  $\langle L_\alpha^{\text{intr}} \rangle$ . The prediction of a power-law profile of Ly $\alpha$  haloes is also supported from the Monte Carlo Ly $\alpha$  RT calculation of the cosmological hydrodynamic simulation (Gronke & Bird 2017), which can also be fitted by a power-law profile with a similar slope.

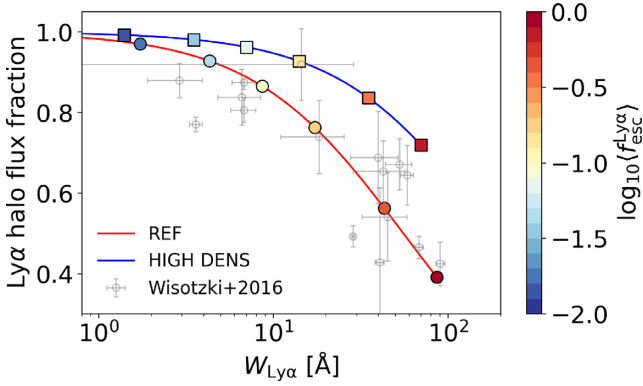
Interestingly, the observation of the surface brightness profiles of Ly $\alpha$  haloes around QSOs by VLT/MUSE by Borisova et al. (2016) reports a shallower power-law profile  $\propto r^{-1.8}$  out to  $\sim 100 \text{ pkpc}$ . This could suggest a large LyC leakage which lowers the surface brightness at inner radius or other Ly $\alpha$  emission mechanism such as fluorescence by the UV photons. Our constrained RT approach can be applied to understand the origin of the Ly $\alpha$  haloes of QSOs (e.g. Ly $\alpha$  scattering versus fluorescence) by extending a joint Ly $\alpha$  emission-absorption analysis (Hennawi & Prochaska 2013) using the QSO-Ly $\alpha$  forest clustering data (Prochaska et al. 2013; Font-Ribera et al. 2013).

The power-law emission tail of Ly $\alpha$  haloes may have an implication for the large-scale clustering of Ly $\alpha$  emission detected by Croft et al. (2016), which will be discussed in Section 6.2.

#### 5.4.2 Dissecting the impacts of CGM/IGM: the clustering and kinematics of the gas around galaxies

Similar to emergent Ly $\alpha$  line profiles, we can examine the details of the impacts of the gas density, kinematics, and UV background of the CGM and IGM on the Ly $\alpha$  haloes. Fig. 11 shows the different surface brightness profiles of Ly $\alpha$  haloes in different gaseous environments: a high density (HIGH DENS), a high inflow velocity (HIGH INFLOW), high velocity dispersion (HIGH VEL. DISP.), and low UV background (LOW UVB) models.

The clustering of neutral hydrogen gas around galaxies increases the surface brightness of the Ly $\alpha$  haloes. The increasing amount of gas increases the probability that Ly $\alpha$  photons are scattered back to lines of sight. For example, the high-density environment



**Figure 12.** Ly $\alpha$  halo flux fraction,  $X_{\text{Ly}\alpha, \text{halo}} = F_{\text{halo}}/(F_{\text{gal}} + F_{\text{halo}})$ , as a function of observed Ly $\alpha$  EW of the emission line of galaxies in the REF (red curve with filled circles) and HIGH DENS model (red curve with filled squares). The points corresponds to  $\langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle = 0.02, 0.05, 0.10, 0.20, 0.50, 1.0$  from left to right. The filled circles are colour-coded with the total Ly $\alpha$  escape fraction,  $\langle f_{\text{esc}}^{\text{Ly}\alpha} \rangle = \langle f_{\text{esc,ISM}}^{\text{Ly}\alpha} \rangle \langle \mathcal{T}_\alpha \rangle$ , as indicated by the colour bar. The data points from the MUSE Hubble Deep Field South by Wisotzki et al. (2016) are shown as open grey points.

with the correlation length of  $r_c = 2 h^{-1}$  cMpc increases the surface brightness profile by about a factor of two. This simultaneously increases the amount of Ly $\alpha$  absorption in that region probed by the background QSO spectra.

Unlike the Ly $\alpha$  line profile, the kinematics of the gas around galaxies only has a secondary role in determining the surface brightness profile of the Ly $\alpha$  haloes. This is because the surface brightness profile is integrated over the frequency, the velocity information is averaged over. There is none the less a small impact on the surface brightness profile. As the inflowing gas has more probability to scatter Ly $\alpha$  photons as the photons is redshifted into the line centre, higher inflow velocity can increase the surface brightness. Higher velocity dispersion similarly increases the probability that Ly $\alpha$  photons are scattered back to our line of sight at the outskirts of Ly $\alpha$  haloes, where the Ly $\alpha$  photons escape otherwise because of the dominating Hubble flow.

The photoionization rate of the UV background impacts the surface brightness of Ly $\alpha$  haloes. A lower photoionization rate increases the neutral fraction of gas, thus more Ly $\alpha$  photons can be scattered back into lines of sight to form larger Ly $\alpha$  haloes.

## 6 IMPLICATIONS

### 6.1 On the origin of Ly $\alpha$ escape and haloes: how do the CGM and IGM affect Ly $\alpha$ radiation in and around galaxies?

We discuss how the ISM, CGM, and IGM affect Ly $\alpha$  escape and Ly $\alpha$  haloes around galaxies. Wisotzki et al. (2016) recently found an anticorrelation in the halo flux fraction–Ly $\alpha$  EW relation. In other words, larger Ly $\alpha$  halo flux fractions are generally found around galaxies that have a lower EW of Ly $\alpha$  emission-line emission coming *directly* from the galaxy. Fig. 12 compares our dust-free ISM + CGM/IGM models with varying ISM escape fractions with observations (see Section 3.3.2 for a description of how we compute the halo flux fraction). Fig. 12 shows that:

(i) The agreement between the observed anticorrelation between Ly $\alpha$  halo flux fraction and Ly $\alpha$  EW and our model suggests that anticorrelation is indeed caused by the spatial ‘redistribution’ of the galaxy’s Ly $\alpha$  flux into extended Ly $\alpha$  haloes via scattering. Higher

H I gas densities more easily scatter the nebular Ly $\alpha$  emission out of the line of sight of an observer, into a more diffuse Ly $\alpha$  fog. In this picture, varying gas density shifts the halo flux fraction–Ly $\alpha$  EW relation diagonally.

Most Ly $\alpha$  RT occurs on interstellar scales (see Section 5.3). If we invoke scattering as the main driver behind the observed anticorrelation between Ly $\alpha$  halo flux fraction and Ly $\alpha$  EW, then this must occur on interstellar scales. Note that ‘interstellar’ scattering here refers to *all* scattering at  $r < r_{\text{min}}$ . This interpretation sharply contrasts with a picture in which Ly $\alpha$  escape is dominated by dust. In the picture, dust does not only reduce the EW, but it also reduces the Ly $\alpha$  flux that escapes into the CGM, and therefore the surface brightness of Ly $\alpha$  haloes. This leads to an interesting implication: if interstellar RT dominates the Ly $\alpha$  scattering process (which it likely does), and if this scattering process provides the physical reason for the observed anticorrelation between Ly $\alpha$  halo flux fraction and Ly $\alpha$  EW, the role of dust on interstellar Ly $\alpha$  is weaker than expected. The H I gas density and kinematics of the ISM is a primary driver in regulating the Ly $\alpha$  escape.

(ii) Our analysis also shows that Ly $\alpha$  scattering in the CGM/IGM cannot be ignored, even at  $z \sim 2-3$ . We expect that the CGM and IGM to contribute to the scatter in the Ly $\alpha$  halo flux fraction–Ly $\alpha$  EW relation (at fixed EW). A higher H I gas density in the CGM/IGM increases the surface brightness of Ly $\alpha$  haloes via scattering. Increasing H I gas density in the CGM/IGM enhances the Ly $\alpha$  halo flux fraction with a small suppression of Ly $\alpha$  EW. This leads to a (almost) vertical shift in the halo flux fraction–Ly $\alpha$  EW relation.

For  $W_{\text{Ly}\alpha} \gtrsim 20$  Å objects (which correspond approximately to narrow-band selected LAEs), the observed scatter in Wisotzki et al. (2016) may be (partially) due to the environmental dependence of Ly $\alpha$  escape fraction and Ly $\alpha$  haloes introduced by the CGM and IGM. This interpretation is consistent with results of Matsuda et al. (2012) – the spatial extent of Ly $\alpha$  haloes around LAEs ( $W_{\text{Ly}\alpha} \gtrsim 20$  Å) depends on the Mpc-scale environment.

The role of the CGM and IGM on the Ly $\alpha$  escape fraction and Ly $\alpha$  haloes through its environmental dependence can be tested observationally. If true, we should see a ‘thicker’ Ly $\alpha$  forest around galaxies with brighter Ly $\alpha$  haloes and a lower Ly $\alpha$  EW. The constrained RT approach provides us with quantitative predictions for the amount of Ly $\alpha$  absorption in the forest (e.g. Fig. 6). Such result can be tested with the MUSE QSO field data, where we expect both the measurements of individual Ly $\alpha$  haloes and the Ly $\alpha$  forest absorption in the vicinity of the Ly $\alpha$  emitting galaxies are possible. Alternatively, we could perform the Ly $\alpha$  forest tomography technique (Lee et al. 2014) in the MUSE Hubble Deep Field South field using bright background LBGs.

This discussion – while speculative – nevertheless clearly demonstrates the effectiveness of a joint Ly $\alpha$  emission-absorption analysis using (integral field) spectroscopic survey of galaxies in QSO fields in shedding new light on the physical origin of Ly $\alpha$  escape and haloes.

### 6.2 Cosmology with Ly $\alpha$ emitting galaxies

#### 6.2.1 Diffuse Ly $\alpha$ emission in the Universe

Our joint analysis with the galaxy-Ly $\alpha$  forest clustering data predicts the power-law emission tail of Ly $\alpha$  haloes. The Ly $\alpha$  haloes may thus contain more Ly $\alpha$  emission at larger scale than the exponential scale length, which is typically fit to the observations of



Ly $\alpha$  haloes. The power-law surface brightness profile of Ly $\alpha$  haloes (equation 28) contains  $\sim 1.5$  times more Ly $\alpha$  emission at radii larger than the exponential scale length ( $25.2 \text{ pkpc} < r_{\perp} < 5 \text{ pMpc}$ ) than the exponential profile of Steidel et al. (2011). Beyond the knee of the exponential profile, after matching the normalization of the power-law profile at 60 pkpc with the exponential profile, the power-law tail at  $> 60 \text{ pkpc}$  contains  $\sim 2\text{--}3$  times more Ly $\alpha$  emission than the exponential surface brightness profile of Ly $\alpha$  haloes. Therefore, the power-law emission tail of Ly $\alpha$  haloes contribute more to the diffuse Ly $\alpha$  emission of the Universe detected by Croft et al. (2016) than estimated using the conventional exponential profile. Although this boosts a contribution from Ly $\alpha$  haloes around star-forming galaxies to the diffuse Ly $\alpha$  emission background, the collective contribution from the Ly $\alpha$  emission from the power-law tails around galaxies still likely fall short below to explain the total Ly $\alpha$  emission of Croft et al. (2016).

### 6.2.2 Ly $\alpha$ RT effect on the large-scale LAE clustering

The non-gravitational bias in the large-scale clustering of LAEs by the Ly $\alpha$  RT (Zheng et al. 2011; Behrens et al. 2018) may complicate a cosmological analysis in HETDEX. In this paper, we showed that the impact of the large-scale gaseous environments around galaxies on Ly $\alpha$  line flux can be estimated from galaxy-Ly $\alpha$  forest cross-correlation data. This means that if the IGM modulates the visibility of Ly $\alpha$  line of galaxies, and consequently the clustering signal of Ly $\alpha$ -selected galaxies, we also expect a correlation between the Ly $\alpha$  EW of emission lines and galaxy-Ly $\alpha$  forest cross-correlation function. For example, the measurement of the LBG-Ly $\alpha$  forest cross-correlation function as a function of different Ly $\alpha$  EW subsamples of LBGs could test the impact of Ly $\alpha$  RT effect on the large-scale clustering of Ly $\alpha$  emitting galaxies.

The large-scale UV background fluctuations, which may also affect the Ly $\alpha$  forest autocorrelation function (Pontzen 2014; Gontcho A Gontcho, Miralda-Escudé & Busca 2014), can perturb the visibility of Ly $\alpha$  lines and would induce an additional non-gravitational contribution (Wyithe & Dijkstra 2011). To circumvent these issues, various joint HETDEX+BOSS analysis among Ly $\alpha$  line profile, galaxy-Ly $\alpha$  forest cross-correlation, auto-correlation function of Ly $\alpha$  emitting galaxies, and Ly $\alpha$  forest auto-correlation function, are important to quantify the astrophysical impact of the CGM/IGM on Ly $\alpha$  line of galaxies, and hence isolate the cosmological contribution in the large-scale LAE clustering from non-gravitational ones. A further study of Ly $\alpha$  RT modelling and Ly $\alpha$  forest using cosmological hydrodynamical simulations of galaxies and the IGM is required to address this issue more carefully.

### 6.3 Caveats

The analysis presented in this paper is at an early stage. Ideally the same galaxy population should be selected for all observations: galaxy-Ly $\alpha$  forest clustering, Ly $\alpha$  escape fraction, and Ly $\alpha$  haloes. The present analysis relied on LBG sample for galaxy-Ly $\alpha$  forest clustering, which has then been assumed to be the same for all galaxies that escape fraction and Ly $\alpha$  haloes are measured from. This clearly introduces uncertainties in our results. For example, for LAEs because the gas overdensity around lower mass haloes is expected to be lower, the surface brightness of the Ly $\alpha$  haloes may be reduced as less photons are scattered back into lines of sight. While we have explored the possible variation by varying the model parameters around the calibrated value, future analysis will

benefit from the measurements of all the three observables from the uniformly selected sample.

While our main approach – constrained Ly $\alpha$  RT – can substantially reduce the number of free parameters, a particular model adopted in this paper still contains some arbitrariness. Some of which, e.g. gas velocity field, can be improved by directly fitting the model to the data by the Markov Chain Monte Carlo method. We have for example neglected the impact of outflow. The outflow can allow more Ly $\alpha$  photons to escape and reduce the surface brightness of Ly $\alpha$  haloes at smaller radii. Although the currently available 2D galaxy-Ly $\alpha$  forest clustering measurement does not show clear outflow signature, once the measurement is improved, the model should be generalized to allow the effect of outflow. Furthermore, the analysis will benefit by replacing the fundamental assumption of the model, i.e. single scattering approximation, with a full Monte Carlo Ly $\alpha$  RT calculation. Although the single scattering approximation seems to match the full Monte Carlo result reasonably well (Dijkstra & Kramer 2012), the full calculation eliminates the need of somewhat arbitrary choice of  $r_{\text{min}}$  in the model (see Appendix C for the assessment of the uncertainty).

## 7 CONCLUSIONS

We have presented a new approach to ‘constrained Ly $\alpha$  RT’ through the CGM and IGM based on a perturbative expansion of the scattering process (Sections 2 and 3). The central idea is to perform Ly $\alpha$  RT calculations through a CGM and IGM whose H I content and kinematics are constrained by the observed cross-correlation between galaxies and Ly $\alpha$  forest absorption lines (see Section 5.1). This enables us to quantify how Ly $\alpha$  propagates around galaxies in realistic, observationally constrained, cosmological environments. We apply this approach to investigate how Ly $\alpha$  escapes from galaxies and their environment. We specifically focused on studying (i) how the CGM/IGM impacts the Ly $\alpha$  line flux, and spectral line profiles that we observe directly from galaxies, and (ii) the mean surface brightness profile of extended Ly $\alpha$  emission produced by scattering in the CGM/IGM. Our analysis, which focused on  $z \sim 2\text{--}3$ , showed that

(i) The CGM and IGM at  $z \sim 2\text{--}3$  transmit  $\langle \mathcal{T}_{\alpha} \rangle \approx 80$  per cent of Ly $\alpha$  line emission escaping from galaxies at a Ly $\alpha$  velocity offset of  $\Delta v_{\text{Ly}\alpha} \approx 300 \text{ km s}^{-1}$  (redward of the line centre). The transmission varies from  $\approx 60$  per cent at  $\Delta v_{\text{Ly}\alpha} \approx 100 \text{ km s}^{-1}$  to  $\approx 90$  per cent at  $\Delta v_{\text{Ly}\alpha} \approx 600 \text{ km s}^{-1}$ . This excess attenuation of Ly $\alpha$  photons by the CGM/IGM is due to the large overdensity of CGM gas relative to that of mean IGM. The wavelength dependence of the IGM transmission curve near the systemic velocity of a galaxy is determined by the gas kinematics of the CGM. Comparing with observational constraints on the Ly $\alpha$  escape fraction at these redshifts (derived for example from Ly $\alpha$ /H $\alpha$  ratios, e.g. Hayes et al. 2010), our model implies that the ISM plays the biggest role in setting the Ly $\alpha$  escape fraction. This confirms the traditional view that the ISM is a primary driver of the Ly $\alpha$  escape. However, our results demonstrate that there is a non-negligible impact of the CGM and IGM on the Ly $\alpha$  line even at  $z \approx 2\text{--}3$ .

(ii) We show that Ly $\alpha$  scattering in the CGM gives rise to Ly $\alpha$  haloes with a power-law emission tail  $\propto r^{-2.4}$  at  $\gtrsim 80 \text{ pkpc}$  extending out to the outskirts of the CGM and beyond ( $\gtrsim 300 \text{ pkpc}$ , see equation 28). This result is robust, and does not depend on interstellar RT effects. This power-law profile differs from the often-assumed exponential profile, for which the surface brightness drops rapidly beyond the exponential scale length. The extended power-law tail is



a result of the extended clustering of cold gas around galaxies, that is required by the galaxy- $\text{Ly}\alpha$  forest clustering data. This gas scatters  $\text{Ly}\alpha$  photons back into our line of sight to form  $\text{Ly}\alpha$  haloes. Deeper observations of  $\text{Ly}\alpha$  haloes probing the outskirts of the CGM will test this picture. Moreover, if we assume that *all*  $\text{Ly}\alpha$  photons produced in the ISM eventually leak out into the CGM and IGM (note that ‘ISM’ here refers to  $r < r_{\min}$ ), then  $\text{Ly}\alpha$  scattering alone can reproduce the observed surface brightness as well as the observed anticorrelation between  $\text{Ly}\alpha$  halo flux fraction and  $\text{Ly}\alpha$  EW (see Section 6 and Fig. 12), surprisingly well. Of course, the 100 per cent  $\text{Ly}\alpha$  leakage into the CGM is at face value unrealistic. However, the surprisingly good match may still imply that  $\text{Ly}\alpha$  scatters abundantly at  $r < r_{\min}$  without being efficiently destroyed by dust (see Section 6), which places interesting constraints on the  $\text{H}\text{I}$  and dust distribution in the ISM (see e.g. Gronke et al. 2017).

(iii) The CGM and IGM cause a preferential suppression of the  $\text{Ly}\alpha$  flux – and therefore by extension  $\text{Ly}\alpha$  EW – of galaxies, and an overall enhancement of the surface brightness of  $\text{Ly}\alpha$  haloes around galaxies that reside in denser parts of the Universe (more precisely, inside cosmic volumes with a higher  $\text{Ly}\alpha$  forest opacity in the background QSO spectra). The impact of the CGM and IGM can then introduce a large-scale environmental dependence of  $\text{Ly}\alpha$  line profiles and  $\text{Ly}\alpha$  haloes of galaxies. Studies of the environmental impact on  $\text{Ly}\alpha$  line profiles and  $\text{Ly}\alpha$  haloes, i.e.  $\text{Ly}\alpha$  emission properties as a function of the large-scale gaseous environment, are important to fully understand the escape mechanism of  $\text{Ly}\alpha$  photons and the physical origin of  $\text{Ly}\alpha$  haloes.

Finally, we would like to stress the tremendous potential of the joint  $\text{Ly}\alpha$  emission – absorption measurements for studying the physics of CGM, reionization, and cosmology:

(i) From the theoretical side, the analysis introduced in this paper can easily be expanded in various ways. For example, one can study the redshift evolution of  $\langle T_{\alpha} \rangle$ ,  $\text{Ly}\alpha$  haloes, and the UV background self-consistently to address the impact of the CGM and UV background on the decline of observed  $\text{Ly}\alpha$  emission line at  $z > 6$  to study the reionization process (Sadoun et al. 2017). Furthermore, the model can easily be extended to e.g. (i) predict the spectra of the spatially scattered  $\text{Ly}\alpha$  haloes, and (ii) to include the contribution of fluorescence radiation around galaxies and QSOs (e.g. Cantalupo et al. 2005; Kollmeier et al. 2010; Mas-Ribas & Dijkstra 2016).

(ii) From the observational side, wide-field imaging campaigns such as the Dark Energy Survey and Hyper Suprime-Cam Subaru Strategic Program will provide new high-redshift QSO target fields. By combining these observations with data obtained with modern integral field spectrographs such as VLT/MUSE and Keck/KCWI (and in the near future *James Webb Space Telescope* and ground-based 30 m telescopes), it becomes possible to provide a cosmic map of both  $\text{Ly}\alpha$  emission and absorption. HETDEX galaxies and BOSS  $\text{Ly}\alpha$  forests will also offer a promising data set. Such maps provide invaluable probes of the detailed galaxy properties and the physical state of the CGM and IGM at  $2 \lesssim z \lesssim 7$ . This approach will shed new light on the physics of CGM, hydrogen and helium reionization, and cosmology.

## ACKNOWLEDGEMENTS

We thank Max Gronke and Mike Anderson for enlightening discussion; Kristian Finlator for pointing out the importance of  $\text{Ly}\alpha$  haloes, which motivated us to examine the properties of  $\text{Ly}\alpha$  haloes in this paper; Rieko Momose for providing us the tabulated surface brightness profile of  $\text{Ly}\alpha$  haloes; and Andrew Pontzen,

Andreu Font-Ribera, and Richard Ellis for reading and the comments on the manuscript. We thank the referee for carefully reading the manuscript and her/his constructive comments. KK acknowledges support from the European Research Council Advanced Grant FP7/669253.

## REFERENCES

- Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, *ApJ*, 584, 45
- Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, *ApJ*, 629, 636
- Becker G. D., Bolton J. S., 2013, *MNRAS*, 436, 1023
- Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, *MNRAS*, 430, 2067
- Behrens C., Niemeyer J., 2013, *A&A*, 556, A5
- Behrens C., Byrohl C., Saito S., Niemeyer J. C., 2018, *A&A*, 614, A31
- Bertschinger E., 1985, *ApJS*, 58, 39
- Bielby R. M. et al., 2017, *MNRAS*, 471, 2174
- Borisova E. et al., 2016, *ApJ*, 831, 39
- Bouwens R. J. et al., 2014, *ApJ*, 793, 115
- Bouwens R. J., Smit R., Labbé I., Franx M., Caruana J., Oesch P., Stefanon M., Rasappu N., 2016, *ApJ*, 831, 176
- Brocklehurst M., 1971, *MNRAS*, 153, 471
- Cai Z. et al., 2017, *ApJ*, 837, 71
- Cantalupo S., Porciani C., Lilly S. J., Miniati F., 2005, *ApJ*, 628, 61
- Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, *Nature*, 506, 63
- Cen R., Miralda-Escudé J., Ostriker J. P., Rauch M., 1994, *ApJ*, 437, L9
- Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, *MNRAS*, 452, 261
- Cooke J., Wolfe A. M., Gawiser E., Prochaska J. X., 2006, *ApJ*, 652, 994
- Crighton N. H. M. et al., 2011, *MNRAS*, 414, 28
- Croft R. A. C. et al., 2016, *MNRAS*, 457, 3541
- Croton D. J. et al., 2006, *MNRAS*, 365, 11
- Davis M., Peebles P. J. E., 1977, *ApJS*, 34, 425
- Dayal P., Maselli A., Ferrara A., 2011, *MNRAS*, 410, 830
- Dekel A. et al., 2009, *Nature*, 457, 451
- Dijkstra M., 2014, *PASA*, 31, e040
- Dijkstra M., 2017, preprint (arXiv:1704.03416)
- Dijkstra M., Kramer R., 2012, *MNRAS*, 424, 1672
- Dijkstra M., Wyithe J. S. B., 2010, *MNRAS*, 408, 352
- Dijkstra M., Lidz A., Wyithe J. S. B., 2007, *MNRAS*, 377, 1175
- Dijkstra M., Mesinger A., Wyithe J. S. B., 2011, *MNRAS*, 414, 2139
- Eldridge J. J., Stanway E. R., Xiao L., McClelland L. A. S., Taylor G., Ng M., Greis S. M. L., Bray J. C., 2017, *PASA*, 34, e058
- Faucher-Giguère C.-A., Kereš D., Dijkstra M., Hernquist L., Zaldarriaga M., 2010, *ApJ*, 725, 633
- Faucher-Giguère C.-A., Feldmann R., Quataert E., Kereš D., Hopkins P. F., Murray N., 2016, *MNRAS*, 461, L32
- Fisher K. B., 1995, *ApJ*, 448, 494
- Font-Ribera A. et al., 2013, *J. Cosmol. Astropart. Phys.*, 5, 018
- Fumagalli M., Prochaska J. X., Kasen D., Dekel A., Ceverino D., Primack J. R., 2011, *MNRAS*, 418, 1796
- Gontcho A Gontcho S., Miralda-Escudé J., Busca N. G., 2014, *MNRAS*, 442, 187
- Grazian A. et al., 2016, *A&A*, 585, A48
- Greig B., Komatsu E., Wyithe J. S. B., 2013, *MNRAS*, 431, 1777
- Gronke M., Bird S., 2017, *ApJ*, 835, 207
- Gronke M., Dijkstra M., 2016, *ApJ*, 826, 14
- Gronke M., Dijkstra M., McCourt M., Oh S. P., 2016, *ApJ*, 833, L26
- Gronke M., Dijkstra M., McCourt M., Oh S. P., 2017, *A&A*, 607, A71
- Haardt F., Madau P., 1996, *ApJ*, 461, 20
- Haardt F., Madau P., 2012, *ApJ*, 746, 125
- Hayes M., 2015, *PASA*, 32, e027
- Hayes M. et al., 2010, *Nature*, 464, 562
- Hennawi J. F., Prochaska J. X., 2013, *ApJ*, 766, 58

Hennawi J. F., Prochaska J. X., Cantalupo S., Arrigoni-Battaia F., 2015, *Science*, 348, 779

Hill G. J. et al., 2008, in Kodama T., Yamada T., Aoki K., eds, ASP Conf. Ser. Vol. 399, Panoramic Views of Galaxy Formation and Evolution, Astron. Soc. Pac., San Francisco. p. 115

Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014, *MNRAS*, 445, 581

Hui L., Gnedin N. Y., Zhang Y., 1997, *ApJ*, 486, 599

Iliev I. T., Shapiro P. R., McDonald P., Mellema G., Pen U.-L., 2008, *MNRAS*, 391, 63

Inoue A. K., Shimizu I., Iwata I., Tanaka M., 2014, *MNRAS*, 442, 1805

Kakiichi K., Dijkstra M., Ciardi B., Graziani L., 2016, *MNRAS*, 463, 4019

Kennicutt R. C., Jr., 1998, *ApJ*, 498, 541

Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, 363, 2

Kollmeier J. A., Zheng Z., Davé R., Gould A., Katz N., Miralda-Escudé J., Weinberg D. H., 2010, *ApJ*, 708, 1048

Krogager J.-K. et al., 2013, *MNRAS*, 433, 3091

Lake E., Zheng Z., Cen R., Sadoun R., Momose R., Ouchi M., 2015, *ApJ*, 806, 46

Laursen P., Sommer-Larsen J., Razoumov A. O., 2011, *ApJ*, 728, 52

Lee K.-G. et al., 2014, *ApJ*, 795, L12

Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2015, *MNRAS*, 446, 3697

Madau P., 1995, *ApJ*, 441, 18

Madau P., Pozzetti L., Dickinson M., 1998, *ApJ*, 498, 106

Mas-Ribas L., Dijkstra M., 2016, *ApJ*, 822, 84

Mas-Ribas L., Dijkstra M., Hennawi J. F., Trenti M., Momose R., Ouchi M., 2017, *ApJ*, 841, 19

Mason C. A., Treu T., Dijkstra M., Mesinger A., Trenti M., Pentericci L., de Barros S., Vanzella E., 2017, *ApJ*, 856, 2

Matsuda Y. et al., 2012, *MNRAS*, 425, 878

McCourt M., Oh S. P., O’Leary R. M., Madigan A.-M., 2016, *MNRAS*, 473, 5407

Meiksin A., 2006, *MNRAS*, 365, 807

Meiksin A., Bolton J. S., Tittle E. R., 2014, *MNRAS*, 445, 2462

Meiksin A., Bolton J. S., Tittle E. R., 2015, *MNRAS*, 453, 899

Mesinger A., Aykutalp A., Vanzella E., Pentericci L., Ferrara A., Dijkstra M., 2015, *MNRAS*, 446, 566

Momose R. et al., 2014, *MNRAS*, 442, 110

Murray N., Quataert E., Thompson T. A., 2005, *ApJ*, 618, 569

Noterdaeme P. et al., 2012, *A&A*, 540, A63

Paresce F., McKee C. F., Bowyer S., 1980, *ApJ*, 240, 387

Pontzen A., 2014, *Phys. Rev. D*, 89, 083010

Prochaska J. X. et al., 2013, *ApJ*, 776, 136

Rahmati A., Schaye J., Bower R. G., Crain R. A., Furlong M., Schaller M., Theuns T., 2015, *MNRAS*, 452, 2034

Rakic O., Schaye J., Steidel C. C., Rudie G. C., 2012, *ApJ*, 751, 94

Reid B. A., White M., 2011, *MNRAS*, 417, 1913

Robertson B. E. et al., 2013, *ApJ*, 768, 71

Rudie G. C. et al., 2012, *ApJ*, 750, 67

Rudie G. C., Steidel C. C., Shapley A. E., Pettini M., 2013, *ApJ*, 769, 146

Sadoun R., Zheng Z., Miralda-Escudé J., 2017, *ApJ*, 839, 44

Santos M. R., 2004, *MNRAS*, 349, 1137

Schaye J., 2001, *ApJ*, 559, 507

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

Shen S., Madau P., Guedes J., Mayer L., Prochaska J. X., Wadsley J., 2013, *ApJ*, 765, 89

Siana B. et al., 2015, *ApJ*, 804, 17

Smith R. E. et al., 2003, *MNRAS*, 341, 1311

Steidel C. C., Erb D. K., Shapley A. E., Pettini M., Reddy N., Bogosavljević M., Rudie G. C., Rakic O., 2010, *ApJ*, 717, 289

Steidel C. C., Bogosavljević M., Shapley A. E., Kollmeier J. A., Reddy N. A., Erb D. K., Pettini M., 2011, *ApJ*, 736, 160

Trainor R. F., Steidel C. C., Strom A. L., Rudie G. C., 2015, *ApJ*, 809, 89

Tummuangpak P., Bielby R. M., Shanks T., Theuns T., Crighton N. H. M., Francke H., Infante L., 2014, *MNRAS*, 442, 2094

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, *MNRAS*, 445, 794

Turner M. L., Schaye J., Crain R. A., Rudie G., Steidel C. C., Strom A., Theuns T., 2017, *MNRAS*, 471, 690

van de Voort F., Schaye J., 2012, *MNRAS*, 423, 2991

Veilleux S., Cecil G., Bland-Hawthorn J., 2005, *ARA&A*, 43, 769

Verhamme A., Schaerer D., Maselli A., 2006, *A&A*, 460, 397

White S. D. M., Frenk C. S., 1991, *ApJ*, 379, 52

Wiese W. L., Fuhr J. R., 2009, *J. Phys. Chem. Reference Data*, 38, 565

Wisotzki L. et al., 2016, *A&A*, 587, A98

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, 43, 861

Worseck G. et al., 2014, *MNRAS*, 445, 1745

Wyithe J. S. B., Dijkstra M., 2011, *MNRAS*, 415, 3929

Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2010, *ApJ*, 716, 574

Zheng Z., Cen R., Trac H., Miralda-Escudé J., 2011, *ApJ*, 726, 38

## APPENDIX A: DERIVATION OF LY $\alpha$ HALO SURFACE BRIGHTNESS

We describe a heuristic derivation of the mean surface brightness of Ly $\alpha$  haloes powered by the scattering of Ly $\alpha$  photons from central galaxies. The total bolometric Ly $\alpha$  luminosity within a comoving radius  $R$  is

$$\langle L_{\alpha}(<R) \rangle = \int_0^R dr \iint L_{\alpha}^{\text{abs}}(r, v_r, N_{\text{HI}}) f(r, v_r, N_{\text{HI}}) dv_r dN_{\text{HI}}, \quad (\text{A1})$$

where  $L_{\alpha}^{\text{abs}}(r, v_r, N_{\text{HI}})$  is a Ly $\alpha$  luminosity of an individual absorber at a comoving distance  $r$  away from the central galaxy, moving with a radial peculiar velocity  $v_r$ , with an H I column density  $N_{\text{HI}}$ , and  $f(r, v_r, N_{\text{HI}}) dr dv_r dN_{\text{HI}}$  is the phase-space distribution function of absorbers around galaxies, i.e. the expected number of absorbers within a phase space volume  $(r, r + dr)$ ,  $(v_r, v_r + dv_r)$ , and  $(N_{\text{HI}}, N_{\text{HI}} + dN_{\text{HI}})$ . The phase-space distribution function can be decomposed into a product of a real-space distribution function,  $f_r(r, N_{\text{HI}}) dr dN_{\text{HI}}$ , and the conditional probability distribution function of the peculiar velocities of absorbers at a given  $r$ ,  $p_v(v_r|r) dv_r$ , leading  $f(r, v_r, N_{\text{HI}}) = f_r(r, N_{\text{HI}}) p_v(v_r|r)$ .

The real-space distribution function is simply given by

$$f_r(r, N_{\text{HI}}) dr dN_{\text{HI}} = \frac{dn_{\text{abs}}}{dN_{\text{HI}}} [1 + \xi(r)] 4\pi r^2 dr dN_{\text{HI}}, \quad (\text{A2})$$

where  $\frac{dn_{\text{abs}}}{dN_{\text{HI}}}$  is the comoving number density of absorbers in a range between  $N_{\text{HI}}$  and  $N_{\text{HI}} + dN_{\text{HI}}$  and  $\xi(r)$  is the real-space correlation function between galaxies and absorbers. Note that  $\frac{dn_{\text{abs}}}{dN_{\text{HI}}}$  is related to the H I density distribution function  $\frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z}$  as (Wolfe, Gawiser & Prochaska 2005)

$$\frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} dN_{\text{HI}} dz = (1 + z)^3 \frac{dn_{\text{abs}}}{dN_{\text{HI}}} \sigma_{\text{abs}} \left| \frac{dl_p}{dz} \right| dN_{\text{HI}} dz. \quad (\text{A3})$$

Furthermore, for the Gaussian streaming model the conditional probability distribution function of the peculiar velocity of absorbers around galaxies is

$$p_v(v_r|r) = \frac{1}{\sqrt{2\pi\sigma_v^2(r)}} \exp \left[ -\frac{(v_r - \langle v_r(r) \rangle)^2}{2\sigma_v^2(r)} \right]. \quad (\text{A4})$$

Therefore, using equation (14) for the Ly  $\alpha$  luminosity of individual absorbers, after some algebra, we obtain

$$\begin{aligned} \langle L_\alpha(< R) \rangle &= \int_0^R dr \int dN_{\text{HI}} \left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}} \frac{dz}{dr} \\ &\times \int \frac{dv_r}{\sqrt{2\pi\sigma_v^2(r)}} \mathcal{L}_\alpha(r, v_r, N_{\text{HI}}) [1 + \xi(r)] \\ &\exp \left[ -\frac{(v_r - \langle v_r(r) \rangle)^2}{2\sigma_v^2(r)} \right], \end{aligned} \quad (\text{A5})$$

where we simplified the expression by introducing an auxiliary quantity,

$$\mathcal{L}_\alpha(r, v_r, N_{\text{HI}}) = \int [1 - e^{-\tau_\alpha(v_{\text{inj}}, N_{\text{HI}})}] L_v^{\text{intr}}(v_e) dv_e. \quad (\text{A6})$$

As the comoving bolometric emissivity (luminosity density) (in units of  $\text{erg s}^{-1} \text{cMpc}^{-3}$ ) is related to total luminosity within comoving radius  $R$ ,  $\int_0^R \langle \varepsilon_\alpha(v, r) \rangle 4\pi r^2 dr = \langle L_\alpha(< R) \rangle$ , we have  $dL_\alpha(< R)/dR = \langle \varepsilon_\alpha(R) \rangle 4\pi R^2$ . Thus, the Ly  $\alpha$  emissivity by the scattered radiation is given by

$$\begin{aligned} \langle \varepsilon_\alpha(r) \rangle &= \frac{1}{4\pi r^2} \int dN_{\text{HI}} \left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}} \frac{dz}{dr} \\ &\times \int \frac{dv_r}{\sqrt{2\pi\sigma_v^2(r)}} \mathcal{L}_\alpha(r, v_r, N_{\text{HI}}) [1 + \xi(r)] \\ &\exp \left[ -\frac{(v_r - \langle v_r(r) \rangle)^2}{2\sigma_v^2(r)} \right], \end{aligned} \quad (\text{A7})$$

which conclude the derivation of equation (15). The mean surface brightness of Ly  $\alpha$  haloes follows immediately by integrating the Ly  $\alpha$  emissivity along a line of sight at each impact parameter.

## APPENDIX B: IMPACT OF PHOTOIONIZATION BY LOCAL IONIZING SOURCES

By taking the Jeans argument by Schaye (2001), the H I column density of an absorber scales as  $N_{\text{HI}} \propto \Gamma^{-1}$  as

$$\begin{aligned} N_{\text{HI}} &\sim 2.7 \times 10^{13} \text{ cm}^2 (1 + \delta)^{3/2} \left( \frac{T}{10^4 \text{ K}} \right)^{-0.26} \left( \frac{\Gamma}{10^{-12} \text{ s}^{-1}} \right)^{-1} \\ &\times \left( \frac{1+z}{4} \right)^{9/2} \left( \frac{\Omega_b h^2}{0.02} \right)^{3/2} \left( \frac{f_g}{0.16} \right)^{1/2}. \end{aligned} \quad (\text{B1})$$

This means that at a fixed density fluctuation  $\delta$ , different photoionization rates give rise to different H I column densities of absorbers. Therefore, the impact of photoionization can be taken into account by rescaling the value of H I column density at a photoionization rate  $\Gamma_{\text{bkg}}$  to a new value of photoionization rate  $\Gamma(r) = \Gamma_{\text{local}}(r) + \Gamma_{\text{bkg}}$  which includes the photoionization rate from local ionizing sources  $\Gamma_{\text{local}}(r)$ . The H I column density of absorbers in the vicinity of galaxies is given by rescaling the column density of absorbers in the average IGM  $N_{\text{HI}}^{\text{bkg}}$ ,

$$N_{\text{HI}} = \left[ \frac{\Gamma_{\text{bkg}}}{\Gamma(r)} \right] N_{\text{HI}}^{\text{bkg}}. \quad (\text{B2})$$

Following Hui et al. (1997), the CDDF around galaxies including the photoionization rate of the local sources is given by rescaling the CDDF at the average IGM,

$$\left. \frac{\partial^2 \mathcal{N}(r)}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}} = \frac{\Gamma(r)}{\Gamma_{\text{bkg}}} \left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}}. \quad (\text{B3})$$

While here we explicitly indicated  $\left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}}$  as the CDDF at the average UV background, the subscript bkg is dropped in the main text. Approximating the CDDF with a power law,

$$\left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}} = A(z) (N_{\text{HI}}^{\text{bkg}})^{-\beta_{\text{eff}}}, \quad (\text{B4})$$

where  $A(z)$  is a constant of proportionality, we obtain the CDDF around galaxies as

$$\begin{aligned} \left. \frac{\partial^2 \mathcal{N}(r)}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}} &= A(z) \left[ \frac{\Gamma(r)}{\Gamma_{\text{bkg}}} \right]^{-\beta_{\text{eff}}+1} N_{\text{HI}}^{-\beta_{\text{eff}}} \\ &= \left[ \frac{\Gamma(r)}{\Gamma_{\text{bkg}}} \right]^{-\beta_{\text{eff}}+1} \left. \frac{\partial^2 \mathcal{N}}{\partial N_{\text{HI}} \partial z} \right|_{\text{bkg}}. \end{aligned} \quad (\text{B5})$$

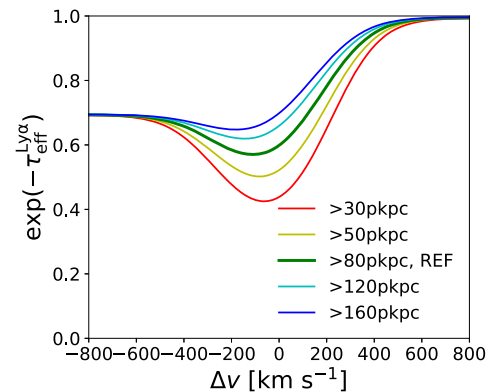
The factor  $\left[ \frac{\Gamma(r)}{\Gamma_{\text{bkg}}} \right]^{-\beta_{\text{eff}}+1}$  indicates the radial suppression of the normalization of the CDDF due to the photoionization by local ionizing sources. Therefore, by comparing with the definition of the real-space correlation function, we can identify the photoionization correction factor  $C^{\text{phot}}(r)$  as

$$C^{\text{phot}}(r) = \left[ \frac{\Gamma(r)}{\Gamma_{\text{bkg}}} \right]^{-\beta_{\text{eff}}+1} = \left[ \frac{\Gamma_{\text{local}}(r)}{\Gamma_{\text{bkg}}} + 1 \right]^{-\beta_{\text{eff}}+1}. \quad (\text{B6})$$

By substituting  $\Gamma_{\text{local}}(r) = \int_{v_{912}}^{\infty} \sigma_{\text{HI}}(v) \frac{L_v(v)}{4\pi[r/(1+z)]^2} \frac{dv}{hv}$ , after some algebra we arrive at equations (23) and (24).

## APPENDIX C: THE INNERMOST RADIUS

Fig. C1 shows the dependence of the Ly  $\alpha$  visibility on the innermost radius  $r_{\text{min}}$  in the REF model, where  $r_{\text{min}} = 30, 50, 80, 120, 160$  pkpc (red, yellow, green, cyan, and blue curves). The virial radius of  $9 \times 10^{11} M_\odot$  halo at  $z = 3$  is  $r_{\text{vir}} \approx 80$  pkpc. The opacity increases with a smaller innermost radius. Our fiducial choice adopted in this paper is  $r_{\text{min}} = r_{\text{vir}}$ . The difference between the recommendation of Laursen et al. (2011),  $r_{\text{min}} = 1.5r_{\text{vir}}$ , with our fiducial choice is about 20 per cent. Using our fiducial ISM Ly  $\alpha$  line profile, the Ly  $\alpha$  transmission is  $\langle \mathcal{T}_\alpha \rangle = 0.76, 0.80, 0.84, 0.86, 0.88$  for  $r_{\text{min}} = 30, 50, 80, 120, 160$  pkpc. The extra scattering between the virial radius of halo (80 pkpc) to the inner region of the model CGM (30 pkpc) is  $\langle \mathcal{T}_\alpha(> 80 \text{ pkpc}) \rangle - \langle \mathcal{T}_\alpha(> 30 \text{ pkpc}) \rangle = 0.08$ . There-



**Figure C1.** Dependence of the CGM/IGM transmission curve on the innermost radius.

fore, the estimated Ly $\alpha$  transmission is therefore robust against the choice of the innermost radius at about 10 per cent level.

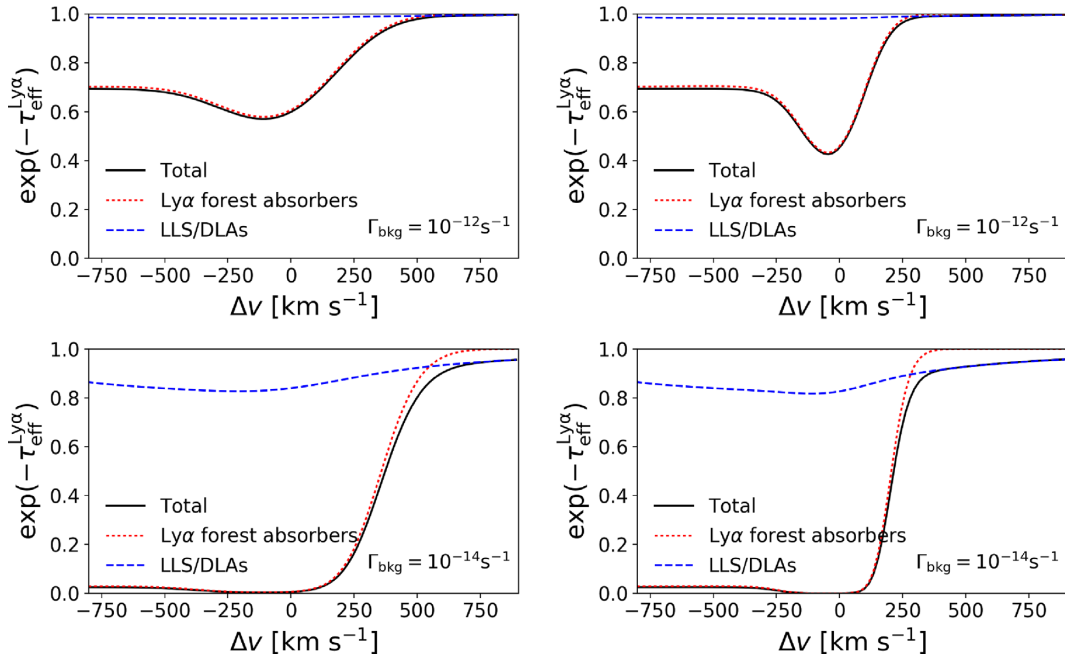
Decreasing the innermost radius to  $r_{\min} = 30$  pkpc increases the opacity by  $\sim 30$  per cent relative to  $r_{\min} = r_{\text{vir}}$ . Laursen et al. (2011) showed that below a virial radius the multiple scatterings affects the formation of Ly $\alpha$  lines; hence,  $e^{-\tau}$  approximation becomes increasingly inaccurate. On this scale, an Ly $\alpha$  RT simulation is required to estimate the impact of the gas around galaxies on the Ly $\alpha$  line profile. As a compromise between the recommendation of Laursen et al. (2011) and the smallest radius probed by Ly $\alpha$  absorption using galaxy–galaxy and galaxy–QSO pairs (below this scale, the properties of gas around galaxies are not well constrained), we choose  $r_{\min} = 80$  pkpc to evaluate the impact of the intergalactic environment ‘constrained by Ly $\alpha$  absorption’ on the Ly $\alpha$  flux from Ly $\alpha$  emitting galaxies.

#### APPENDIX D: RELATIVE CONTRIBUTION OF ABSORBERS

Fig. D1 shows the contribution of different absorbers and different velocity dispersion to the red damping wing. At  $\Gamma_{\text{bkg}} = 10^{-12} \text{ s}^{-1}$ , the velocity dispersion of infalling low column density Ly $\alpha$  for-

est absorbers dominate the formation of red damping wing. The smoothness of the damping wing depends on the velocity dispersion of the CGM: a higher  $\sigma_v$  smear out the damping wing. Note that the velocity dispersion parameter is the most poorly constrained parameter in our model due to the lack of tabulated 2D effective optical depth map. For example, decreasing to  $\sigma_v = 100 \text{ km s}^{-1}$ , the damping wing extends only to  $\Delta v \approx 250 \text{ km s}^{-1}$ , more consistent with the results of Laursen et al. (2011) [see also Dijkstra et al. (2007) where the effect of velocity dispersion is ignored].

Our finding is consistent with Kakiichi et al. (2016), where we find that at  $z \sim 7$  LLS and DLAs dominate the contribution to the red damping wing due to small-scale absorbers (e.g. W2 model). By taking a value of  $\Gamma_{\text{bkg}} = 10^{-14} \text{ s}^{-1}$ , Fig. D1 indeed shows that the contribution from LLSs and DLAs dominate the red damping wing opacity at  $\Delta v > 250 - 500 \text{ km s}^{-1}$ , while it is negligible at  $\Gamma_{\text{bkg}} = 10^{-12} \text{ s}^{-1}$ . This is because at  $\Gamma_{\text{bkg}} = 10^{-12} \text{ s}^{-1}$ , although the effective optical depth at large  $\Delta v$  is dominated by LLS/DLAs, its absolute value is very small,  $\tau_{\text{eff}}^{\text{Ly}\alpha} \ll 1$ ; hence, it is negligible in  $\exp(-\tau_{\text{eff}}^{\text{Ly}\alpha})$ . As  $\tau_{\text{eff}}^{\text{Ly}\alpha} \propto \Gamma_{\text{bkg}}^{-1/2}$ , at a lower photoionization rate  $\Gamma_{\text{bkg}} = 10^{-14} \text{ s}^{-1}$ , the effective optical depth becomes  $\tau_{\text{eff}}^{\text{Ly}\alpha} \gtrsim \mathcal{O}(1)$ , making an appreciable contribution to  $\exp(-\tau_{\text{eff}}^{\text{Ly}\alpha})$ .



**Figure D1.** The CGM/IGM transmission curves including all absorbers (black solid,  $10^{12} < N_{\text{HI}}/\text{cm}^{-2} < 10^{21.55}$ ), only Ly $\alpha$  forest absorbers (red dotted,  $10^{12} < N_{\text{HI}}/\text{cm}^{-2} < 10^{17}$ ), and only LLS and DLAs (blue dashed,  $10^{17} < N_{\text{HI}}/\text{cm}^{-2} < 10^{21.55}$ ). Left-hand panels: models with the velocity dispersion  $\sigma_v = 200 \text{ km s}^{-1}$  and  $\Gamma_{\text{bkg}} = 10^{-12} \text{ s}^{-1}$  (top) or  $\Gamma_{\text{bkg}} = 10^{-14} \text{ s}^{-1}$  (bottom). Right-hand panels: model with the velocity dispersion  $\sigma_v = 100 \text{ km s}^{-1}$  and  $\Gamma_{\text{bkg}} = 10^{-12} \text{ s}^{-1}$  (top) or  $\Gamma_{\text{bkg}} = 10^{-14} \text{ s}^{-1}$  (bottom). All the other parameters are the same as REF model. The figure shows the contribution of the different types of absorbers to the CGM/IGM transmission curves.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.