# The Very Large Telescope Lyman-Break Galaxy Redshift Survey – IV. Gas and galaxies at $z \sim 3$ in observations and simulations

P. Tummuangpak,[1]★ R. M. Bielby,[1] T. Shanks,[1] T. Theuns,[2,3] N. H. M. Crighton,[4] H. Francke[5] and L. Infante[5]

[1]*Department of Physics, University of Durham, South Road, Durham DH1 3LE*
[2]*Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE*
[3]*Department of Physics, University of Antwerp, Campus Groenenborger, Groenenborgerlaan 171, B-2020 Antwerp, Belgium*
[4]*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*
[5]*Departamento de Astronomía y Astrofísica, Pontificia Universidad Catolica de Chile, Casilla 306, Santiago 22, Chile*

## ABSTRACT

We use a combination of observations and simulation to study the relationship between star-forming galaxies and the intergalactic medium at $z \approx 3$. The observed star-forming galaxy sample is based on spectroscopic redshift data taken from a combination of Very Large Telescope (VLT) Lyman-break galaxy (LBG) Redshift Survey (VLRS) data and Keck Low-Resolution Imaging Spectrometer (LRIS) observations in fields centred on bright background quasi-stellar objects (QSOs), whilst the simulation data is taken from the Galaxies–Intergalactic Medium Interaction Calculation (GIMIC). In the simulation, we find that the dominant peculiar velocities are in the form of large-scale coherent motions of gas and galaxies. Gravitational infall of galaxies towards one another is also seen, consistent with expectations from linear theory. At smaller scales, the root-mean-square (RMS) peculiar velocities in the simulation overpredict the difference between the simulated real- and $z$-space galaxy correlation functions. Peculiar velocity pairs with separations smaller than $1\,h^{-1}$ Mpc have a smaller dispersion and explain the $z$-space correlation function better. The Ly$\alpha$ auto- and cross-correlation functions in the GIMIC simulation appear to show infall smaller than implied by the expected $\beta_{\rm Ly\alpha} \approx 1.3$ (McDonald et al.). There is a possibility that the reduced infall may be due to the galaxy-wide outflows implemented in the simulation. The main challenge in comparing these simulated results with the observed Keck + VLRS correlation functions comes from the presence of velocity errors for the observed LBGs, which dominate at $\lesssim 1\,h^{-1}$ Mpc scales. When these are taken into account, the observed LBG correlation functions are well matched by the high amplitude of clustering, shown by higher mass ($M_* > 10^9\,\mathrm{M_\odot}$) galaxies in the simulation. The simulated cross-correlation function shows similar neutral gas densities around galaxies to those seen in the observations. The simulated and observed Ly$\alpha$ $z$-space autocorrelation functions again agree better with each other than with the $\beta_{\rm Ly\alpha} \approx 1.3$ infall model. Our overall conclusion is that, at least in the simulation, gas and galaxy peculiar velocities are generally towards the low end of expectation. Finally, little direct evidence is seen in either simulation or observations for high transmission near galaxies due to feedback, in agreement with previous results.

**Key words:** galaxies: high-redshift – intergalactic medium.

## 1 INTRODUCTION

The effect of feedback via supernovae and active galactic nucleus (AGN) driven winds is thought to be a key factor in the pro-cess of galaxy formation and evolution. Cosmological models of galaxy formation require efficient injection of feedback from su-pernovae (SNe) and AGNs to regulate the star formation activity and thus replicate the observed galaxy stellar mass function (e.g. White & Rees 1978; White & Frenk 1991). Similarly, cosmologi-cal simulations, for example those of Springel & Hernquist (2003), Schaye et al. (2010) and Scannapieco et al. (2012), have shown that

★ E-mail: pimpunyawat.tummuangpak@durham.ac.uk

supernova feedback is fundamental to recreating cosmic star-formation history. It is also evident that simulations lacking some sort of feedback struggle to reproduce realistic disc galaxies (e.g. Weil, Eke & Efstathiou 1998; Schaye et al. 2010; McCarthy et al. 2012b; Scannapieco et al. 2012) and that powerful galactic winds are required to produce the observed metal enrichment of the intergalactic medium (IGM) (e.g. Cen & Ostriker 1999; Theuns et al. 2002; Aguirre et al. 2005; Oppenheimer & Davé 2006).

In terms of observing the effects of feedback at high redshift, Adelberger et al. (2003, hereafter A03) presented the cross-correlation between $z \sim 3$ galaxies and the IGM (as traced by quasar sightlines) and claimed an observed lack of absorbing gas within $\sim 0.5 \, h^{-1}$ Mpc. They interpreted this as evidence of strong galactic winds removing H I gas from the vicinity of these star-forming galaxies. The work was based on Keck High Resolution Echelle Spectrometer (HiRES; $R \sim 40\,000$) spectra of eight background quasars at $z \approx 3$ combined with 431 Lyman-break galaxies (LBGs) from the survey of Steidel et al. (2003). Following the results of A03, Adelberger et al. (2005a, hereafter A05) updated the result with greater numbers of galaxies, this time centred at $z \sim 2$. Based on this new sample, A05 found an increase in Ly$\alpha$ absorption down to scales of $r \sim 0.5 \, h^{-1}$ Mpc of LBG positions, with no evidence for H I gas having been removed from the vicinity of these galaxies. Indeed, Crighton et al. (2011) surmised that the cross-correlation at such small scales would likely be affected by uncertainties in galaxy redshifts in the A03 data. It is therefore still unclear to what extent galactic winds have an effect on this probe of galaxy surroundings.

In addition to the above evidence for gas outflows, gas infall down to galaxy scales is also predicted in models of galaxy formation (e.g. White & Rees 1978; Kereš et al. 2005; Dekel & Birnboim 2006; Kereš et al. 2009; Dekel et al. 2009; van de Voort et al. 2012). Gas inflow is expected to be coherent down to the virial radius of a massive galaxy ($\approx$140 kpc), below which scale the situation is more complicated due to shocks and the gas pressure becoming more important. Gas flow infall into galaxies along filaments is also expected in secular models of galaxy formation, where the gas accretion rate may not be simply dictated by merging rates in a hierarchical model (Dekel et al. 2009). Rakic et al. (2012) presented a study of the galaxy–Ly$\alpha$ cross-correlation at $z \approx 2.4$ using 15 fields of the Keck Baryonic Structure Survey (KBSS). They saw fingers-of-God on sub-500 kpc scales and evidence for infall on $\sim$8 Mpc scales.

In order to constrain models of galaxy formation, it is imperative to provide extensive observations of the IGM via hydrogen and metal absorption lines and thus identify and probe the infall and outflow processes. As such, we are undertaking a large galaxy survey centred on distant bright quasars in the form of the Very Large Telescope (VLT) LBG Redshift Survey (VLRS). Bielby et al. (2011) presented the first stage of the galaxy survey, comprising $\approx$1000 $z \sim 3$ galaxies within $\sim$30 arcmin of $z > 3$ quasars. Using this sample, Crighton et al. (2011) performed a cross-correlation analysis between the galaxy positions and the Ly$\alpha$ forest of the available quasar spectra in the fields, finding increased absorption within $\sim 5 \, h^{-1}$ Mpc of galaxy positions. This result was consistent with the results of A03 and A05, but lacked the galaxy numbers to probe the $\sim 0.5 \, h^{-1}$ Mpc scales at which A03 claimed to see the effects of galaxy winds. Since then, the VLRS has been extended to incorporate $\sim$2000 LBGs within nine separate fields containing bright $z > 3$ quasars (Bielby et al. 2013), comparable in number to the only other equivalent surveys at this redshift (e.g. Rakic et al. 2012; Rudie et al. 2012).

A number of authors have provided complimentary analysis of such galaxy–gas correlations at $z \sim 3$ using smoothed particle hydrodynamical simulations (e.g. Croft et al. 2002; Kollmeier et al. 2003; Bruscoli et al. 2003; Desjacques et al. 2004; Desjacques, Haehnelt & Nusser 2006; Rakic et al. 2013). Partly prompted by the first survey of LBGs in bright quasar fields, Croft et al. (2002) and Kollmeier et al. (2003) both investigated possible explanations for the enhancement in the gas profile around LBGs reported by A03 and the distribution of gas around high-redshift galaxies in general, using smoothed-particle hydrodynamics (SPH) simulations. Croft et al. (2002) found that the absorption profiles around high-redshift galaxies increase monotonically with decreasing distance from the galaxies in their simulations. Similarly, Kollmeier et al. (2003) presented results consistent with Croft et al. (2002), showing that, based on their SPH simulations, photoionization cannot explain the observed reduction in absorption presented by A03.

More recently, Rakic et al. (2013) used the OverWhelmingly Large Simulation (OWLS), comparing analysis of OWLS with their own observational results (Rakic et al. 2012). As with previous simulation work, the authors find a continuous increase in absorption with decreasing distance from a galaxy, consistent with their observations. They go on to analyse the 2D H I Ly$\alpha$ absorption profile and claim a good match between their observations and the simulation, with the gas distribution on scales of $\sim$8 Mpc being consistent with large-scale gas infall into the potential wells occupied by galaxies.

In this article, we update the work of Crighton et al. (2011), adding the galaxy redshifts of Bielby et al. (2011) and also Steidel et al. (2003) in conjunction with the available high-resolution quasar spectra in these survey fields. This work thus combines the higher galaxy sampling rate of the Steidel et al. (2003) survey with the wide fields of the VLRS and provides a galaxy sample that can probe the full range of scales from a few hundred kpc to tens of Mpc. This large range of scales is imperative for distinguishing between models of gas inflow and outflow in 2D galaxy–Ly$\alpha$ cross-correlation analysis. In addition to extending the previous observational results, we also incorporate a hydrodynamical simulation, the Galaxies–Intergalactic Medium Interaction Calculation (GIMIC: Crain et al. 2009), into our analysis in order to interpret the observations.

This article is organized as follows. Observational data from the VLT LBG Redshift Survey and Keck LBG observations of Steidel et al. (2003) are described in Section 2. Section 3 describes the GIMIC simulations. The simulated galaxy clustering results and their interpretation are shown in Section 4, while the galaxy–IGM cross-correlation is presented in Section 5. Section 6 presents an analysis of the Ly$\alpha$ autocorrelation in both the observations and the simulation. Our discussion and conclusions are presented in Sections 7 and 8 respectively.

Throughout this work, we adopt a cosmology consistent with the GIMIC simulation (and hence the Millennium simulation, Springel et al. 2005). This corresponds to $\{\Omega_m, \Omega_\Lambda, \Omega_b, n_s, \sigma_8, H_0, h\} = \{0.25, 0.75, 0.045, 1, 0.9, 100, 0.73\}$. As we are working in both real and redshift space in this article, it is prudent to note the conventions on coordinates that we use here. For real-space separations between two points we use $r$, whilst in redshift space we use $s$. Where a plot shows results in both real and redshift space (i.e. where we show simulation results), we denote the distance axis with $r$. For the observed data, all distances are of course measured in redshift space and so separations are denoted by $s$ in any plots primarily showing observational data. We denote the transverse and line-of-sight coordinates as $\sigma$ and $\pi$ respectively, regardless of whether these

are in real or redshift space. All coordinates are given in comoving coordinates in this article unless stated otherwise.

## 2 OBSERVATIONS

In this work, we use a combination of spectroscopically identified $z \sim 3$ star-forming galaxies and high-resolution spectral observations of the Ly$\alpha$ forest of $z \gtrsim 3$ quasars. The galaxy data are a combination of the VLRS data presented by Bielby et al. (2011, 2013), and the publicly available Keck LBG data presented by Steidel et al. (2003). These two data sets are based on different observing strategies, whereby the VLRS offers coverage across large fields of view, whilst the Keck sample covers relatively small separations ($\lesssim 8$–10 Mpc) with higher sampling rates of the galaxy population. The quasar spectra with which we trace the distribution of H I within the fields have all been obtained from archival VLT Ultraviolet and Visual Echelle Spectrograph (UVES) and Keck HiRES observations. In this section, we give details of all data and reduction processes used for the quasar spectra.

### 2.1 LBG observations

The VLRS currently provides $\sim$2000 spectroscopic galaxy redshifts within nine fields centred on $z \gtrsim 3$ quasars (Bielby et al. 2011, 2013). The redshifts were obtained using the VLT VIsible MultiObject Spectrograph (VIMOS) instrument (Le Fèvre et al. 2003) with the LR_Blue grism, giving a resolution of $R \sim 180$ and velocity accuracies of $\sigma_v \approx 350\,\mathrm{km\,s^{-1}}$. In total, the survey covers an area of $\sim$2.6 deg$^2$ and provides galaxy data in the foreground of the following nine high-redshift quasars: Q0042$-$2627 ($z = 3.29$), J0124+0044 ($z = 3.84$), Q0301$-$0035 ($z = 3.23$), HE0940$-$1050 ($z = 3.05$), J1201+0116 ($z = 3.23$), PKS2126$-$158 ($z = 3.28$), Q2231+0015 ($z = 3.02$), Q2348$-$011 ($z = 3.02$) and Q2359+0653 ($z = 3.23$). The spectroscopic galaxy sample is predominantly limited to $R < 25$ (Vega), although a number of fainter galaxies ($R < 25.5$ Vega) are present in the sample, where slit allocation during the VIMOS observations could be optimized by their inclusion.

The LBG redshifts were identified using Ly$\alpha$ emission lines and interstellar medium (ISM) absorption lines, where visible. For both the Ly$\alpha$ and ISM features, it is necessary to correct the measured redshift for intrinsic velocity effects, due to these features being affected by outflowing gas (e.g. A03: Steidel et al. 2010). As such, the VLRS galaxy redshifts have been corrected according to the prescription given by Steidel et al. (2010).

The Keck survey provides a sample of $\sim$940 LBGs observed using the Keck Low-Resolution Imaging Spectrometer (LRIS) instrument (Oke et al. 1995). The quasars from six (Q0201+1120, Q0256$-$0000, Q0302$-$0019, B0933+2854, Q2233+1341 and Q1422+2309) out of the 17 Keck fields are available to us through the public archive; taking only those galaxies in fields around these six Keck quasars, the number of LBGs is reduced to 308. The Keck LBGs are limited to $\mathcal{R} = 25.5$ (AB).

### 2.2 Quasar data

We have analysed publicly available archival spectroscopy for 16 quasars in the redshift range $2.9 \lesssim z \lesssim 3.6$, with an additional quasar spectrum provided by our own X-Shooter observations to make a total of 17 quasar sightlines. The publicly available data are all high-resolution ($R \gtrsim 30\,000$), high signal-to-noise (S/N $\gtrsim 20$) spectra observed using either the UVES instrument (Dekker et al. 2000) on the VLT or the HiRES instrument (Vogt et al. 1994) on the Keck

telescope. Full details of the reduction of UVES and HiRES quasar spectra for 11 of the quasars used here are provided by Crighton et al. (2011). The remaining six spectra were all observed with the Keck HiRES instrument and were reduced following an identical method to that used for the two Keck quasars of Crighton et al. (2011), using the MAKEE package.[1] Briefly, this encompassed basic flat-fielding and bias subtraction, followed by the use of SPIM2 to splice the echelle orders and combine individual observations. This involved producing template spectra constructed by combining the individual observations, masking bad regions of the CCDs and rescaling. A template was applied to rescale the original observations. We divide out the continuum for each individual observation, then multiply this normalized flux by a continuum fit to the template. After scaling each order of each observation individually, we combined them to get the final spectrum.

In addition to the publicly available quasar spectra, we also include a spectrum from our own observations using the X-Shooter instrument (Vernet et al. 2011) on the VLT for the quasar Q2359+0653. These data were reduced using the X-Shooter pipeline package; see Bielby et al. (in preparation) for details. The full list of quasars used in this study is provided in Table 1.

## 3 GIMIC SIMULATIONS

### 3.1 Overview

We simulate both Ly$\alpha$ spectra and galaxies for comparison with observational data using a hydrodynamical cosmological simulation. Our main aims are to study the real- and redshift-space auto- and cross-correlation functions. We wish to ascertain whether we can detect the effects of peculiar velocities, in order to understand more about gas outflow and infall around galaxies, for (a) LBG–LBG pairs (b) Ly$\alpha$–Ly$\alpha$ pairs and (c) the LBG–Ly$\alpha$ forest. The results will then be used to interpret the observable 1D and 2D correlation functions $\xi(r)$ and $\xi(\sigma, \pi)$ in terms of both simulation and observational results. As outlined earlier, $\sigma$ denotes the distance transverse to the line of sight, $\pi$ denotes the line-of-sight distance and $r$ is the (real-space) vector combination of the two coordinates, thus $r = \sqrt{\sigma^2 + \pi^2}$. When working in redshift space we use $s$ in place of $r$.

We use the GIMIC simulation, which is a cosmological hydrodynamical resimulation of selected volumes of the Millennium simulation (Springel et al. 2005). GIMIC is designed to overcome the issues in simulating large cosmological volumes ($L \gtrsim 100\,h^{-1}$ Mpc) at high resolution ($m_{\mathrm{gas}} \lesssim 10^7 h^{-1}\,\mathrm{M_\odot}$) to $z = 0$ by taking a number of smaller regions with 'zoomed' initial conditions (Frenk et al. 1996; Power et al. 2003; Navarro et al. 2004). These individual regions each have approximate radii of $18\,h^{-1}$ Mpc, outside which the remainder of the Millennium simulation volume is modelled with collisionless particles at much lower resolution.

GIMIC was run using the TreePM SPH code GADGET3, which is an update of the GADGET2 code (Springel 2005). The cosmological parameters adopted were $\Omega_{\mathrm{m}} = 0.25$, $\Omega_{\lambda} = 0.75$, $\Omega_{\mathrm{b}} = 0.045$, $h_0 = 100\,h\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, $h = 0.73$, $\sigma_8 = 0.9$ and $n_{\mathrm{s}} = 1$ (where $n_{\mathrm{s}}$ is the spectral index of the primordial power spectrum).

Radiation cooling and stellar evolution were implemented as described in Wiersma et al. (2009), whilst star formation was handled as described by Schaye & Dalla Vecchia (2008) and supernova

---

[1] www2.keck.hawaii.edu/inst/common/makeewww

**Table 1.** List of quasars used in this study.

| Quasar | RA | Dec. | $z$ | Mag. | Instrument |
|---|---|---|---|---|---|
| | J2000 | | | | |
| Q2359+0653 | 00:01:40.6 | +07:09:54 | 3.23 | $V = 18.5$ | X-Shooter |
| Q0042−2627 | 00:44:33.9 | −26:11:19 | 3.289 | $B = 18.5$ | HIRES |
| WHO91 0043−265 | 00:45:30.5 | −26:17:09 | 3.44 | $R = 18.3$ | HIRES |
| J0124+0044 | 01:24:03.8 | +00:44:32 | 3.83 | $g = 19.2$ | UVES |
| Q0201+1120 | 02:03:46.7 | +11:34:45 | 3.610 | $G = 20.1$ | HIRES |
| Q0256−0000 | 02:59:05.6 | +00:11:22 | 3.364 | $G = 18.2$ | HIRES |
| Q0301−0035 | 03:03:41.0 | −00:23:22 | 3.230 | $g = 17.6$ | HIRES |
| Q0302−0019 | 03:04:49.9 | −00:08:13 | 3.281 | $G = 17.8$ | HIRES |
| B0933+2845 | 09:33:37.2 | +28:45:32 | 3.428 | $G = 17.5$ | HIRES |
| HE0940−1050 | 09:42:53.5 | −11:04:25 | 3.06 | $B = 17.2$ | UVES |
| J1201+0116 | 12:01:44.4 | +01:16:11 | 3.233 | $g = 17.7$ | HIRES |
| Q1422+2309 | 14:24:38.1 | +22:56:01 | 3.620 | $G = 16.5$ | HIRES |
| Q2129−1602 | 21:29:04.9 | −16:02:49 | 2.90 | $R = 19.2$ | HIRES |
| PKS2126−158 | 21:29:12.2 | −15:38:40 | 3.268 | $V = 17.3$ | UVES |
| Q2231+0015 | 22:34:08.9 | +00:00:01 | 3.02 | $r = 17.3$ | UVES |
| Q2233+1341 | 22:36:27.2 | +13:57:13 | 3.209 | $G = 20.0$ | HIRES |
| Q2348−011 | 23:50:57.9 | −00:52:10 | 3.023 | $r = 18.7$ | UVES |

feedback was implemented following the prescription of Dalla Vecchia & Schaye (2008).

The GIMIC simulations are particularly well suited to the study of $\sim L^*$ galaxies. As shown in Crain et al. (2009), the implementation of efficient (but energetically feasible) feedback from SNe largely prevents overcooling on the mass scale of $L^*$ galaxies and is key to the reproduction of the observed X-ray scaling relation presented in that study. Indeed, GIMIC accurately reproduces the rotation speeds and star-formation efficiencies of $z = 0$ disc galaxies for $10^9 \lesssim M < 10^{10.5}\,\mathrm{M_\odot}$, although galaxies with $M_* \gtrsim 10^{11}\,\mathrm{M_\odot}$ do still suffer from some overcooling (McCarthy et al. 2012b). Moreover, Font et al. (2011) demonstrated that $L^*$ galaxies in GIMIC exhibit satellite luminosity functions and stellar spheroid surface brightness distributions that are comparable to those of the Milky Way and M31, whilst McCarthy et al. (2012a) further demonstrated that this correspondence also extends to their global structure and kinematics.

In terms of reproducing the Ly$\alpha$ forest, Theuns et al. (1998) conducted simulations across a range of resolutions (i.e. gas particle masses) in order to evaluate the effect of resolution on such studies. They found convergence of the mean effective optical depth (at $z = 3$) in their SPH simulations at gas particle masses of $\lesssim 1.4 \times 10^8\,h^{-1}\,\mathrm{M_\odot}$, whilst column density distributions were found to be consistent given gas particle masses of $\lesssim 1.8 \times 10^7\,h^{-1}\,\mathrm{M_\odot}$. Both of these limits are significantly higher than the GIMIC gas particle mass of $1.45 \times 10^6\,h^{-1}\,\mathrm{M_\odot}$ (Crain et al. 2009), indicating that resolution effects are not an issue for our work in terms of the Ly$\alpha$ forest. In terms of the selected dark matter (DM) haloes, the dark matter particle masses in GIMIC are $6.6 \times 10^6\,h^{-1}\,\mathrm{M_\odot}$, which is $\gtrsim 2$ orders of magnitude lower than any halo mass we will be considering in this study.

In this work, we focus on the Ly$\alpha$ forest, i.e. $N_{\mathrm{H\,I}} \lesssim 10^{17}\,\mathrm{cm^{-2}}$. In this regime, the gas is optically thin, such that radiative transfer implementations such as that of Altay et al. (2011) are not necessary.

An area of interest for this study is the effect of supernova (SN) feedback on the local environment of galaxies. GIMIC contains an implementation of SN feedback based on the generation of winds as follows. First, after a delay corresponding to the maximum lifetime of a star that undergo core-collapse SNe, newly formed star particles impart a randomly directed 600 km s$^{-1}$ kick to, on aver-

age, $\eta = 4$ of its neighbours. Here, $\eta$ is the mass loading (defined as $\eta \equiv \dot{m}_{\mathrm{wind}}/\dot{m}_*$) and its value for GIMIC was chosen to match the global star-formation rate density (SFRD) to observational data. The 600 km s$^{-1}$ initial kick is not equivalent to measured outflow velocities, given that it is a 'launch' velocity and is not necessarily what observations measure. In addition, the particles that receive this wind kick are never decoupled from the hydrodynamical calculations, as is done in e.g. Springel & Hernquist 2003; Oppenheimer & Davé 2008; Kereš et al. 2009; Hirschmann et al. 2013, and so they are subject to significant deceleration as they travel into either the host galaxy disc or halo. We also note that the above simulations specifically direct the applied wind kicks perpendicular to the plane of the host galaxy, as opposed to the randomly oriented wind kicks used in GIMIC. Ultimately, the lack of decoupling and the isotropic nature of the wind kicks in GIMIC means that the values of the launch velocity and mass loading used in GIMIC are necessarily higher than in the above studies. We note that the wind launch velocity used in GIMIC is consistent with the higher end of the Ly$\alpha$ wind velocities reported by Pettini et al. (2001) and Shapley et al. (2003), lending the value some legitimacy.

In the work presented here, we use the '0$\sigma$' GIMIC region, which is identified as having a mean density at $z = 1.5$ equal to the mean density of the Universe at that epoch. In addition, we use only one snapshot of this region, chosen to be at a redshift of $z = 3.1$ in order to provide a suitable comparison to our $z \sim 3$ observed population of star-forming galaxies. All the analysis is limited to a sphere of radius $16\,h^{-1}$ Mpc in order to negate the effects of particles being 'moved' out of the analysis region when moved to redshift space. Given a limiting radius of $16\,h^{-1}$ Mpc, the same number of $M_* > 10^8\,h^{-1}\,\mathrm{M_\odot}$ galaxies are present in the region regardless of whether redshift-space distortions (RSD) are applied or not.

### 3.2 Simulated galaxy population

#### 3.2.1 Identifying the galaxy population

The galaxy population is identified in the simulation, based on first identifying the dark matter haloes using a friends-of-friends (FoF: Davis et al. 1985) algorithm. A group-finding algorithm then locates the nearest dark matter halo for each baryonic (gas or star) particle
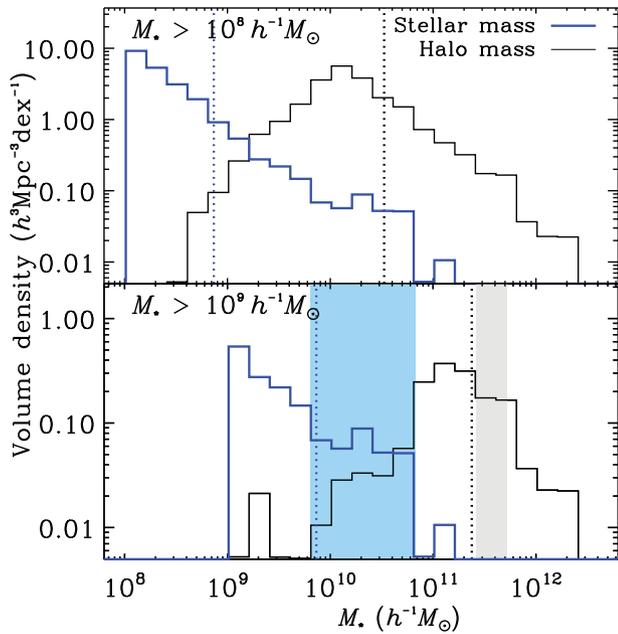
**Figure 1.** Distribution of total halo and galaxy stellar masses for the two GIMIC galaxy selections, $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ (top) and $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ (bottom). The blue histograms in each panel show the number of galaxies as a function of stellar mass, whilst the black histograms show the number of galaxies as a function of total halo mass. The dotted vertical lines show the mean halo mass, $M_{\mathrm{halo}} = 10^{10.5} \, h^{-1} \, \mathrm{M_\odot}$ and $M_{\mathrm{halo}} = 10^{11.4} \, h^{-1} \, \mathrm{M_\odot}$ for the low- and high-mass cuts respectively. The shaded light blue region in the lower panel shows the observed $1\sigma$ range in stellar masses of $z \approx 3$ LBGs from Shapley et al. (2005). See the online paper for a colour version of this figure.

and identifies the particle with this halo. The SUBFIND algorithm (Springel et al. 2001; Dolag et al. 2009) is then used to identify self-bound substructures within the haloes, with which star particles are associated and defined as galaxies.

We use cuts in stellar mass to define our simulated galaxy samples. In the first instance, we take galaxies with stellar masses of $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$. This is intended as a large sample that is not representative of the $z \sim 3$ population sampled by present observations, but acts as a comparison data set for a second more representative sample. Taking our limiting radius within the GIMIC volume of $16 \, h^{-1}$ Mpc radius, this low-mass cut gives a sample of 4070 galaxies from the snapshot at $z = 3.06$ in the $0\sigma$ density region. The distributions of galaxy stellar mass (blue histogram) and host halo mass (black histogram) for this sample are shown for reference in the top panel of Fig. 1. The mean galaxy stellar mass is $M_* = 10^{8.9} \, h^{-1} \, \mathrm{M_\odot}$ (blue vertical dashed line), whilst the mean host halo mass is $M_{\mathrm{halo}} = 10^{10.5} h^{-1} \, h^{-1} \, \mathrm{M_\odot}$ (black vertical dashed line).

With our second simulated galaxy sample, we aim to mimic the observed LBG samples more closely and specifically to reproduce the observed clustering. The $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ cut used above provides a simulated galaxy sample with a clustering signal well matched to the observed clustering of LBGs (see Section 4). Bielby et al. (2013) present measurements of the clustering of the VLRS spectroscopic $z \sim 3$ galaxy sample, estimating a clustering length of $r_0 = 3.83 \pm 0.24 \, h^{-1}$ Mpc and typical halo mass of $10^{11.57 \pm 0.15} \, h^{-1} \, \mathrm{M_\odot}$. Similarly, Adelberger et al. (2005b) measure $r_0 = 4.0 \pm 0.6 \, h^{-1}$ Mpc and halo mass of $10^{11.5 \pm 0.3} \, h^{-1} \, \mathrm{M_\odot}$ for a comparable sample of $z \sim 3$ LBGs. We thus vary the stellar mass constraints on the galaxy selection to match these clustering/mean

halo mass results (where the total masses for the GIMIC galaxies are available from the SUBFIND algorithm). We show the clustering results in Section 4, whilst the resulting stellar and halo mass distributions are shown in the lower panel of Fig. 1. We find that a stellar mass cut of $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ reproduces the observed clustering well and gives a mean halo mass for the simulated galaxies of $M_{\mathrm{halo}} = 10^{11.4} \, h^{-1} \, \mathrm{M_\odot}$, marginally lower than the observed samples but consistent at the $\sim 1\sigma$ level.

The mean of the galaxy stellar masses is $M_* = 10^{9.9} \, h^{-1} \, \mathrm{M_\odot}$ (blue vertical dotted line). This $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ cut gives a sample of 287 simulated galaxies within $16 \, h^{-1}$ Mpc of the centre of the GIMIC volume, equating to a space density of $\rho_{\mathrm{g}} \sim 5 \times 10^{-3} \, h^3 \, \mathrm{Mpc}^{-3}$ (for comparison, Adelberger et al. 2005b measure a space density of $\rho_{\mathrm{g}} = 4 \pm 2 \times 10^{-3} \, h^3 \, \mathrm{Mpc}^{-3}$ for the Keck LBG sample).

The cyan shaded region in the lower panel of Fig. 1 shows the standard deviation range around the mean galaxy stellar mass derived from the observations of Shapley et al. (2005) (i.e. $M_* = 10^{10.32} \, h^{-1} \, \mathrm{M_\odot}$, with a standard deviation of $\sigma_{\log(M^*)} = 0.51$ dex). The galaxy stellar mass of the GIMIC selection overlaps the range of the observed galaxies, but extends further to lower stellar masses (i.e. $M_* < 10^{9.5} \, h^{-1} \, \mathrm{M_\odot}$). We note that the Shapley et al. (2005) result is based on $K_s$ observations and that 23 per cent of their UV selected sample is not included in the stellar mass distribution due to not being detected in the $K_s$ observations. This bias against LBGs that are fainter in the $K_s$ band means that the Shapley et al. (2005) stellar mass distribution lacks some of the lower-mass population, but is unlikely to explain the entire discrepancy between the GIMIC stellar masses and the observed mean galaxy stellar mass. Further to this, Crain et al. (2009) calculate the galaxy stellar mass functions from the GIMIC simulation suite and compare them with observations at $z = 2$, showing that the simulated galaxy mass functions have a significantly steeper slope at $M_* \lesssim 9$–10. They surmise that this reflects a reduction in the efficiency of SN feedback in the simulation for low-mass galaxies.

We also note that the simulation was performed with a relatively high value for $\sigma_8$ (a value of $\sigma_8 = 0.9$, which originated from a combined analysis of the Two-degree-Field Galaxy Redshift Survey (2dFGRS) and Three-Year Wilkinson Microwave Anisotropy Probe (WMAP3): Springel et al. 2005) when compared with the present observed constraints ($\sigma_8 = 0.83 \pm 0.01$: Planck Collaboration et al. 2013) and so, for a given mean halo mass (and galaxy stellar mass), we would expect a higher clustering amplitude from the simulation compared with the observations. This is indeed seen, as although, the mean halo mass and mean galaxy stellar mass are lower in the simulated sample than the observed samples, the clustering amplitude ($r_0 = 4.16 \, h^{-1}$ Mpc) is marginally higher than the observed $r_0$ values from both Bielby et al. (2013) and A05.

All combined, the GIMIC $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ simulated galaxies provide a population that is consistent with the observed LBG population in number density and clustering, although the $M_*$ profiles extend to somewhat lower stellar masses than observed (at least in $K$-band detected samples).

### 3.2.2 Velocity field of the simulated galaxies

The distribution of $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxies in real (black asterisks) and redshift space (squares) is shown in Fig. 2. Throughout this work, we use the $x$ and $y$ coordinates within the simulation as the transverse to the line-of-sight coordinates and $z$ as the line-of-sight coordinate, in either real or redshift space. Fig. 2 illustrates the
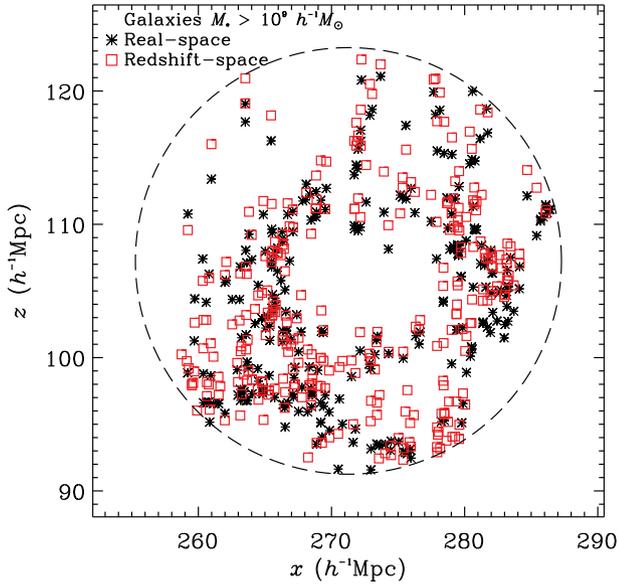
**Figure 2.** The distribution of $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ simulated galaxies in the $x$–$z$ plane in real (black asterisks) and redshift space (open squares), where the $z$-direction is the redshift/line-of-sight dimension in this work. The dashed circle shows the volume limit that we place on the simulation data, given by a radius of $16 \, h^{-1}$ Mpc from the volume centre.

measured positional shifts in the $z$-direction given by the peculiar velocities of the galaxies within the simulation. It is evident from this plot that there is an overall large-scale 'bulk' motion directed in the positive redshift direction due to the motion of the zoomed region with respect to the full $500 \, h^{-1}$ Mpc Millennium volume. Measuring the distribution of the galaxy velocities, we find an average velocity $\langle v \rangle = 93 \, \mathrm{km \, s^{-1}}$ with a standard deviation of $128 \, \mathrm{km \, s^{-1}}$ for the $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxy sample and $\langle v \rangle = +94 \, \mathrm{km \, s^{-1}}$ with a standard deviation of $125 \, \mathrm{km \, s^{-1}}$ for the $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ galaxy sample.

We show the pairwise velocity ($\langle w_z^2 \rangle^{1/2}$) distributions (solid black histograms) of galaxies in Fig. 3 (where $w_z$ is the line-of-sight velocity difference between two objects). For the $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxy samples, we find $\langle w_z^2 \rangle^{1/2} = 172 \, \mathrm{km \, s^{-1}}$ and $\langle w_z^2 \rangle^{1/2} = 176 \, \mathrm{km \, s^{-1}}$, respectively. The dashed histograms show the distribution for only those pairs within $1 \, h^{-1}$ Mpc of each other, thus isolating the intrahalo velocity dispersion and excluding the effect of the halo–halo velocity dispersion. This is important when considering the effect of the velocity dispersion on the galaxy–galaxy clustering measurement. The standard deviations of the pairwise velocities for pairs within $1 \, h^{-1}$ Mpc are 104 and $142 \, \mathrm{km \, s^{-1}}$ for $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxies, respectively. None of these standard deviations includes redshift uncertainties, due to measurement errors that affect the observed galaxy redshifts.

### 3.3 Simulating the Lyα forest spectra

We generate spectra along the $z$-direction through the GIMIC volume. The sightlines were extracted using SPECWIZARD.[2] The transmission is given by $T = \mathrm{e}^{-\tau}$, where $\tau$ is the optical depth along

---

[2] Developed by J. Schaye, C. Booth and T. Theuns; see Theuns et al. (1998) for details.
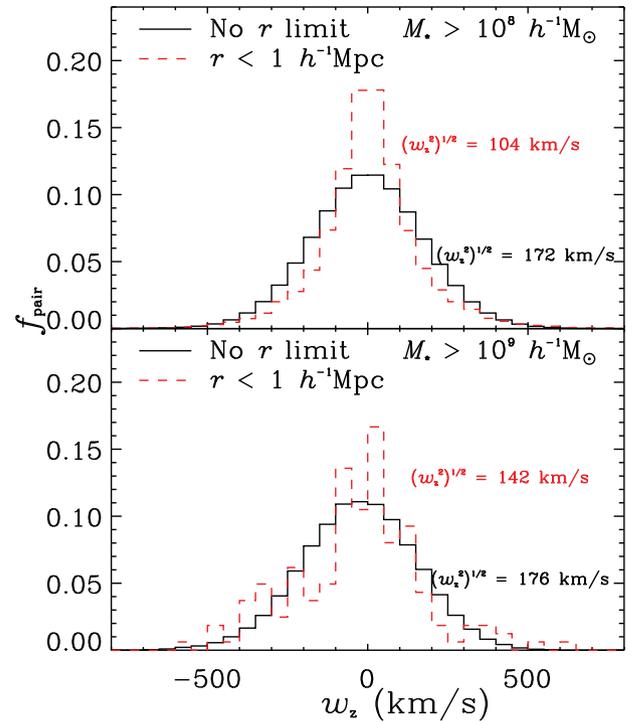


**Figure 3.** The distribution of pairwise velocities ($w_z$, solid histograms) for the GIMIC galaxy samples. The top panel shows the distribution for the $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ galaxy cut and the lower panel that for the $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ cut. Given that the effect of pairwise velocities will be dominant at small scales (i.e. $\lesssim 1 \, h^{-1}$ Mpc), we also show the distributions in each case for only those pairs separated by $r < 1 \, h^{-1}$ Mpc (dashed histograms in both panels). The resulting RMS pairwise velocities are indicated in each case and the separation limit gives smaller values of the RMS pairwise velocity in both cases.

the line of sight. We use a spectral resolution full width at half-maximum (FWHM) of $7 \, \mathrm{km \, s^{-1}}$ to convolve each spectrum, a signal-to-noise ratio of 50 per pixel and pixels of width $2.8 \, \mathrm{km \, s^{-1}}$, which are typical values for our UVES and HIRES quasar spectra. The sightlines were generated parallel to the $z$-axis with random $x$ and $y$ positions. We constructed 200 sightlines, with each sightline being constrained not to extend beyond $16 \, h^{-1}$ Mpc from the centre of the GIMIC volume in order to avoid any edge effects in terms of the gas extent (see Fig. 4). The average transmission, $\bar{T}_r$, for real space is 0.69, while $\bar{T}_z$ for redshift space is 0.72. This difference is likely due to an infall of saturated absorption lines towards each other in redshift space, which results in an overall increase in the measured transmission. This will cause the average transmissivity over the full spectrum to increase in redshift space, as seen. Some hint of this effect can be seen in Fig. 5, in which we show a number of examples of the flux from each sightline compared in real (black lines) and redshift space (dashed lines). These values for the mean transmission at $z \sim 3$ are consistent with the observed values at the $\sim 1 - 2\sigma$ level: for example, McDonald et al. (2000) measure a value of $\bar{T}(z = 3) = 0.684 \pm 0.023$, whilst measurements of the effective optical depth by Faucher-Giguère et al. (2008) give $\bar{T}(z = 3) = 0.680 \pm 0.020$.

Using SPECWIZARD, we calculate the optical depth weighted line-of-sight (LOS) peculiar velocities for each pixel in our 200 spectra. The distribution of the peculiar velocities is given in Fig. 6. As with the galaxy population, the gas traced by the simulated spectra shows the bulk motion in the positive $z$-direction, with a mean
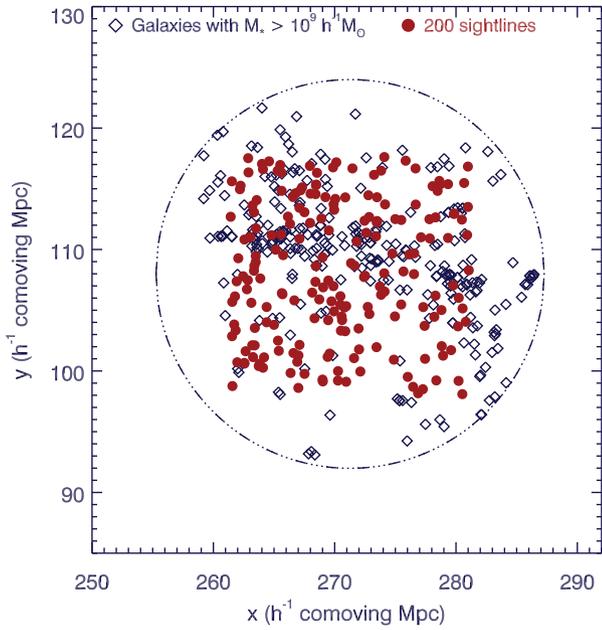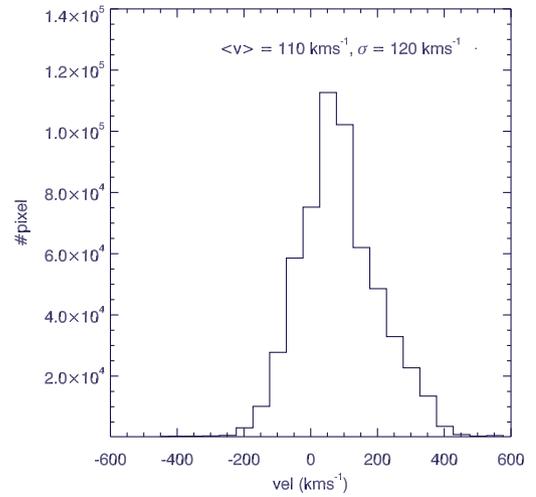
**Figure 6.** The distribution of LOS optical depth weighted peculiar velocities measured within each pixel in each of the GIMIC simulated spectra. This illustrates the underlying dynamics present in the spectra. The LOS peculiar velocity distribution shows a mean peculiar velocity of $\langle v \rangle = 110 \, \mathrm{km \, s^{-1}}$ with a standard deviation of $120 \, \mathrm{km \, s^{-1}}$.
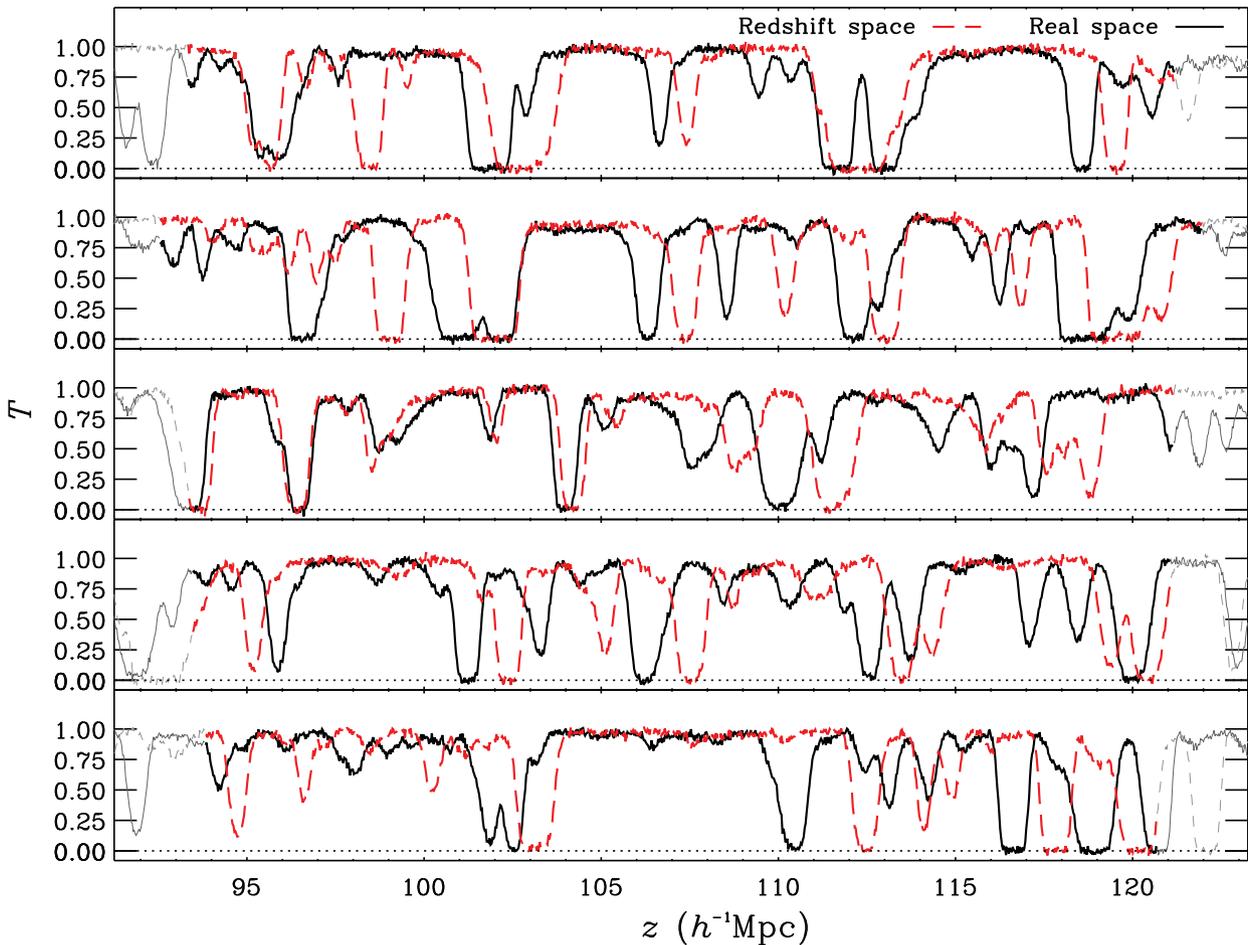
**Figure 4.** The position of $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxies (diamonds) and 200 Ly$\alpha$ sightlines (circles) projected on to the $x$–$y$ plane (i.e. equivalent to an on-sky projection).

**Figure 5.** Examples of absorption spectra from the simulated quasar sightlines. The black solid profiles show $T = e^{-\tau}$ in real space, whilst the dashed line shows $T$ in redshift space. A cut is imposed on the simulated spectra at $r = 16 \, h^{-1}$ Mpc from the centre of the simulation volume. Sections of the spectra that lie outside this sphere are shown in light grey.
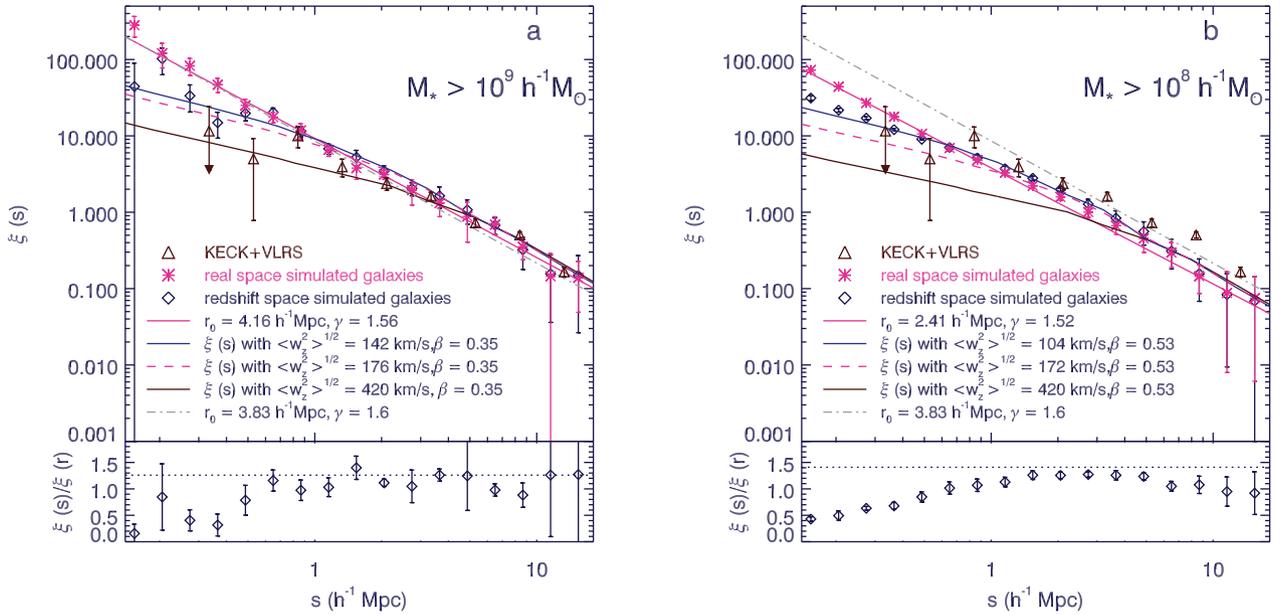
**Figure 7.** (a) Galaxy autocorrelation functions for 287 simulated $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ galaxies compared with the observed LBGs, with the GIMIC real-space result shown by asterisks, the GIMIC redshift-space result shown by blue diamonds and the observed result of Bielby et al. (2013) given by brown triangles. The pink curve shows the power-law fit to the GIMIC real-space clustering, whilst the grey dot–dashed line shows the real-space clustering derived from the observed sample. The pink dashed line is the predicted $\xi(s)$, assuming a real-space fit with $\langle w_z^2 \rangle^{1/2} = 176$ km s$^{-1}$ and $\beta = 0.35$. The blue solid line is the same, except with a $r < 1 \, h^{-1}$ Mpc pairwise dispersion of $\langle w_z^2 \rangle^{1/2} = 142$ km s$^{-1}$. The brown line is the RSD model with velocity errors added to allow comparison with Keck+VLRS LBG $\xi(s)$. (b) The same for 4070 simulated $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxies with real-space fit $\gamma = 1.52$, $r_0 = 2.41 \, h^{-1}$ Mpc. The $\xi(s)$ predictions now assume the appropriate pairwise velocity dispersion of $\langle w_z^2 \rangle^{1/2} = 172$ km s$^{-1}$ (pink dashed line) and $\langle w_z^2 \rangle^{1/2} = 104$ km s$^{-1}$ (blue solid line). The bottom panels present $\xi(s)/\xi(r)$ with jackknife error bars. The dotted line represents the predicted Kaiser boost with (a) $\beta_{\mathrm{gal}} = 0.35$, giving $\xi(s)/\xi(r) = 1.26$ for $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ galaxies, and (b) $\beta_{\mathrm{gal}} = 0.53$, giving $\xi(s)/\xi(r) = 1.41$ for $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxies. See the online paper for a colour version of this figure.

peculiar velocity of $\langle v \rangle = 110$ km s$^{-1}$ and a standard deviation of 120 km s$^{-1}$. The standard deviation of the gas peculiar velocity is comparable to that measured for the galaxy samples ($\approx$125–130 km s$^{-1}$).

## 4 GALAXY CLUSTERING

### 4.1 1D correlation function

#### 4.1.1 Estimator

Bielby et al. (2013) presented a clustering analysis of the LBG data used in this study (combining the VLRS and Keck data). In this section, we compare the observed galaxy clustering presented by Bielby et al. (2013) with results obtained using the galaxy population within the GIMIC simulation. In so doing, we may confirm how representative the GIMIC galaxy population is of the observed $z \approx 3$ LBG population in terms of intrinsic clustering properties and the effects of the galaxy velocity field on galaxy clustering.

We calculate the real- and redshift-space functions, $\xi(r)$ and $\xi(s)$, of the GIMIC $z = 3.06$ galaxy samples using the Davis & Peebles (1983) estimator:

$$\xi(r) = \frac{N_R}{N_G} \frac{\langle DD(r) \rangle}{\langle DR(r) \rangle} - 1, \tag{1}$$

where $\langle DD(r) \rangle$ is the average number of galaxy–galaxy pairs and $\langle DR(r) \rangle$ is the number of galaxy–random pairs at the separation $r$. The factor $N_R/N_G$ is the ratio of the number of random to number of data points.

We estimate the errors in the autocorrelation results using jackknife estimates based on splitting the simulation into equal-volume

octants and excluding each octant in turn to create eight jackknife realizations of the data. The correlation functions are then fitted using a power law of the form

$$\xi(r) = \left( \frac{r}{r_0} \right)^{-\gamma}, \tag{2}$$

where $\gamma$ is the slope of clustering $\xi(r)$ and $r_0$ is the clustering length.

#### 4.1.2 Simulated real-space galaxy correlations

Fig. 7 shows the results for the simulated galaxy–galaxy correlation function with (a) $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ and (b) $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ simulated galaxies. The diamonds show results from galaxies in redshift space, while the pink asterisks show results from galaxies in real space. The integral constraint, $\mathcal{I}$, is included in the data in order to compensate for the effect of the limited field sizes (as described in Bielby et al. 2013). The estimated integral constraints are $\mathcal{I} = 0.21$ and $\mathcal{I} = 0.11$ for $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxies, respectively. The pink lines represent power-law fits to the real-space correlation function based on equation (2). The power-law parameters for the fits to the clustering are given in Table 2. These power-law results give good fits to the real-space clustering results and there is little sign of a double power law or two-halo break in the clustering for either of the samples. However, we note that, in $z \sim 3$ galaxies, the break between the one-halo and two-halo terms is measured to be at $\sim$0.1 arcmin (Hildebrandt et al. 2009), which corresponds to $\approx$0.14 $h^{-1}$ Mpc at $z = 3$. Any break is therefore expected to be at scales smaller than those that we consider in Fig. 7, scales at which we have little sensitivity with which to probe for any possible break.

**Table 2.** Results for the power-law fits to the 1D galaxy autocorrelation functions.

| Sample | $r_0$ ($h^{-1}$ Mpc) | $\gamma$ | Bias | $\beta_{\mathrm{gal}}$ |
|---|---|---|---|---|
| GIMIC $M_* > 10^8\,h^{-1}\,\mathrm{M}_\odot$ | $2.41 \pm 0.24$ | $1.52 \pm 0.10$ | $1.85 \pm 0.12$ | $0.35 \pm 0.04$ |
| GIMIC $M_* > 10^9\,h^{-1}\,\mathrm{M}_\odot$ | $4.16 \pm 1.16$ | $1.56 \pm 0.26$ | $2.80 \pm 0.18$ | $0.23 \pm 0.08$ |
| VLRS (Bielby et al. 2013) | $3.83 \pm 0.24$ | $1.60 \pm 0.09$ | $2.59 \pm 0.13$ | — |

### 4.1.3 Simulated $\xi(s)/\xi(r)$ and infall

In the lower panels of Fig. 7 we show the ratio between real- and redshift-space clustering results from the simulation in order to highlight the signatures of RSD in the redshift-space correlation function. Here the errors are again constructed from the jackknife realizations. At scales above $r \sim 1.5$–$2\,h^{-1}$ Mpc, we see the effects of dynamical infall, which acts to boost the clustering signal in the redshift-space measurement by $\xi(s)/\xi(r) \sim 1.2$–$1.4$. From linear theory (Kaiser 1987; Hamilton 1992), we expect to see a 'Kaiser boost' given by

$$\xi(s) = \left(1 + \frac{2}{3}\beta_{\mathrm{gal}} + \frac{1}{5}\beta_{\mathrm{gal}}^2\right)\xi(r), \tag{3}$$

where $\beta_{\mathrm{gal}}$ is the dynamical infall parameter. For galaxies, $\beta_{\mathrm{gal}} \approx \Omega^{0.6}/b$, where $b$ is the linear galaxy bias and is given by $b = \sqrt{\xi_{\mathrm{gal}}/\xi_{\mathrm{DM}}}$ (here $\xi_{\mathrm{gal}}$ is the galaxy clustering and $\xi_{\mathrm{DM}}$ is the dark matter clustering in real space, Kaiser 1987). At $z \approx 3$, we proceed via the volume-averaged clustering amplitude, $\bar{\bar{\xi}}(8)$, to evaluate both galaxy and dark matter clustering and derive the bias: see equations (17) and (18) of Bielby et al. (2013).

Assuming the power law fitted to $\xi(r)$ for the set of $M_* > 10^8\,h^{-1}\,\mathrm{M}_\odot$ galaxies, we find $\bar{\bar{\xi}}_{\mathrm{g}}(8) = 0.33 \pm 0.02$, giving $b = 1.85 \pm 0.12$ and $\beta_{\mathrm{gal}} = 0.53 \pm 0.07$. At separations of $1 < r < 8\,h^{-1}$ Mpc, we find a mean amplitude ratio of $1.26 \pm 0.03$, which equates to an infall parameter of $0.35 \pm 0.04$. This is lower by $\approx 2.5\sigma$ than the estimate based on the bias. For the $M_* > 10^9\,h^{-1}\,\mathrm{M}_\odot$ simulated galaxy case, the above power-law parameters fitted to $\xi(r)$ give $\bar{\bar{\xi}}_{\mathrm{g}}(8) = 0.75 \pm 0.05$ which, with $\bar{\bar{\xi}}_{\mathrm{DM}}(8) = 0.088$, gives bias $b = 2.80 \pm 0.18$. Taking $\Omega_{\mathrm{m}}(z = 3.0) = 0.98$ gives $\beta_{\mathrm{gal}} = 0.35 \pm 0.02$. The measured Kaiser boost from $\xi(s)/\xi(r)$ is $1.21 \pm 0.06$, which equates to an infall parameter (based on equation 3) of $\beta_{\mathrm{gal}} = 0.23 \pm 0.08$, consistent with what we would expect from the bias at the $\approx 1.5\sigma$ level.

Overall, for both samples we find that the measurements based on the $\xi(s)/\xi(r)(r)$ Kaiser boost appear to result in marginally lower values of $\beta$ than would be expected from the linear theory prediction based on $\beta = \Omega^{0.6}/b$, but only at a $\sim 1$–$2\sigma$ level.

### 4.1.4 Simulated galaxy correlations and velocity dispersion

At smaller separations ($r < 1\,h^{-1}$ Mpc) for both high- and low-mass simulated galaxies, the galaxy–galaxy $\xi(s)$ in redshift space has a lower amplitude than $\xi(r)$. This turn-over of the real-space correlation function is the result of $z$-space smoothing due to the pairwise velocity dispersion, $\langle w_z^2 \rangle^{1/2}$. We model the effects of the pairwise velocity dispersion on the clustering results using a Gaussian profile to the velocity dispersion, following previous work (e.g. Hawkins et al. 2003; da Ângela et al. 2005):

$$f(w_z) = \frac{1}{\sqrt{2\pi}\langle w_z^2\rangle^{1/2}}\exp\left(-0.5\frac{|w_z|^2}{\langle w_z^2\rangle^{1/2}}\right). \tag{4}$$

Using the pairwise velocity dispersions derived from Fig. 3 (i.e. $\langle w_z^2 \rangle^{1/2} = 176\,\mathrm{km\,s^{-1}}$ and $\langle w_z^2 \rangle^{1/2} = 172\,\mathrm{km\,s^{-1}}$ for the $M_* >$

$10^9\,h^{-1}\,\mathrm{M}_\odot$ and $M_* > 10^8\,h^{-1}\,\mathrm{M}_\odot$ samples respectively – pink dashed lines in both panels), we find that the reduction of the real-space clustering at small scales is overpredicted compared with the measurements of $\xi(s)$. As illustrated in Fig. 3, however, we note that the measured pairwise velocity dispersion is separation-dependent. The discrepancy is therefore likely the result of the effect of small-scale peculiar motions on the clustering function, being dominated by galaxies within $\sim 1\,h^{-1}$ Mpc of each other, whereas the initial pairwise velocity histogram presented in Fig. 3 includes pairwise velocities between galaxies across all separation scales within the simulation. If we thus limit the histogram of pairwise velocities to only those pairs within $1\,h^{-1}$ Mpc of each other (dashed histograms in Fig. 3), we retrieve pairwise velocity dispersions of $\langle w_z^2 \rangle^{1/2} = 142\,\mathrm{km\,s^{-1}}$ and $\langle w_z^2 \rangle^{1/2} = 104\,\mathrm{km\,s^{-1}}$ for the $M_* > 10^9\,h^{-1}\,\mathrm{M}_\odot$ and $M_* > 10^8\,h^{-1}\,\mathrm{M}_\odot$ samples, respectively. Using these values in the RSD model, we find improved agreement between the model (solid blue line in Fig. 7) and the galaxy autocorrelation function measured from the GIMIC simulations. Ultimately, the appropriate velocity dispersion for modelling the RSD effects on the galaxy clustering is the velocity dispersion present within groups, whilst the peculiar velocity measured from the simple histogram case included the imprint of the velocity dispersion of galaxy groups as well as the dispersion within groups. Taking the histogram of only pairs of galaxies within $\sim 1\,h^{-1}$ Mpc of each other effectively measures the intragroup peculiar velocities. We conclude that $\xi(s)$ is better described on sub-Mpc scales with the intragroup velocity dispersion appropriate for these scales.

### 4.2 Simulated and observed correlation functions compared

Bielby et al. (2013) report the best-fitting scale-length and slope for the observed Keck + VLRS LBG–LBG semi-projected $w_{\mathrm{p}}(\sigma)$ for the data as $r_0 = 3.83 \pm 0.24\,h^{-1}$ Mpc, with a slope of $\gamma = 1.60 \pm 0.09$. Within the reported errors, the clustering of our $M_* > 10^9\,h^{-1}\,\mathrm{M}_\odot$ sample reproduces the observed survey clustering very well in terms of both clustering length and slope. As would be expected, the $M_* > 10^8\,h^{-1}\,\mathrm{M}_\odot$ sample gives a somewhat lower clustering length than the observational data, but does at least have a consistent slope within the quoted errors.

We now apply the measured $\langle w_z^2 \rangle^{1/2}$ from the observations of Bielby et al. (2013) to our correlation functions measured from GIMIC. Bielby et al. (2013) measured $\langle w_z^2 \rangle^{1/2} = 420\,\mathrm{km\,s^{-1}}$, which includes both the intrinsic velocity dispersion and the velocity errors in measuring the galaxy redshifts. The measured $\xi(s)$ from Bielby et al. (2013) is shown in Fig. 7 (brown triangles) and a model based on the GIMIC $\xi(r)$ combined with the observational $\langle w_z^2 \rangle^{1/2} = 420\,\mathrm{km\,s^{-1}}$ is given by the brown solid line. By introducing the observationally measured pairwise velocity errors to the GIMIC $M_* > 10^9\,h^{-1}\,\mathrm{M}_\odot$ result, we find that the GIMIC clustering measurement reproduces the measured LBG clustering well.

### 4.3 2D correlation function

We now turn to the 2D galaxy autocorrelation functions, in order to investigate the impact of galaxy velocities on clustering

measurements within the simulation further. In the 2D correlation function, $\xi(\sigma, \pi)$, we parametrize the line-of-sight separation between two galaxies by $\pi$ and the transverse separation by $\sigma$. We calculate $\xi(\sigma, \pi)$ using the same methods as used for the 1D correlation functions and with the same samples.

### 4.3.1 Simulations

Figs 8 and 9 show the 2D galaxy autocorrelation function, $\xi(\sigma, \pi)$, for $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ simulated galaxies
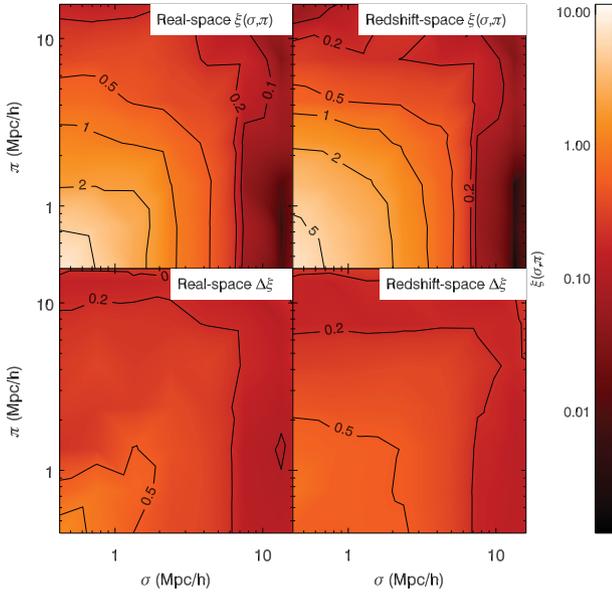


**Figure 8.** The 2D autocorrelation function $\xi(\sigma, \pi)$ results based on the simulated $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ galaxies. The top panels show $\xi(\sigma, \pi)$ measured in real space (left panel) and redshift space (right panel), with a clear shift in the contours in the line-of-sight ($\pi$) direction at small scales, showing the effect of peculiar velocities. Large-scale bulk motions are also in evidence via the flattening of the $\xi = 0.2$ contour at $\pi \sim 10 \, h^{-1}$ Mpc. The lower panels show the error contours over the same scales.
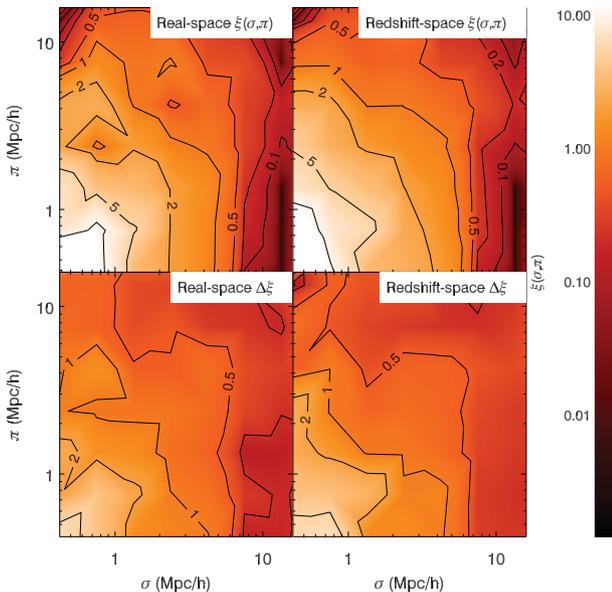


**Figure 9.** As in Fig. 8, but for the GIMIC $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxy sample.

respectively (both with the integral constraint added). In both cases, the top left panel shows the real-space measurement and the top right panel shows the redshift-space measurement. The bottom panels show the respective error contours for the $\xi(\sigma, \pi)$ measurements.

Taking the $M_* > 10^8 \, h^{-1} \, \mathrm{M_\odot}$ results first, the effects of the RSD are clearly visible in the top panels of Fig. 8, where the redshift-space $\xi(\sigma, \pi)$ contours are more extended at scales of $\lesssim 4 \, h^{-1}$ Mpc, whilst being flattened at scales of $\gtrsim 4 \, h^{-1}$ Mpc in comparison with the real-space result. In terms of the latter, the shift in position of the $\xi = 0.5$ and $\xi = 0.2$ contours from the left to right panels is clear evidence of the Kaiser boost.

We now fit this $\xi(\sigma, \pi)$ result with a model based on incorporating the infall parameter, $\beta$, and convolving this with the velocity dispersion (e.g. Hawkins et al. 2003; da Ângela et al. 2005):

$$\xi(\sigma, \pi) = \int_{-\infty}^{\infty} \xi'(\sigma, \pi - w_z(1+z)/H(z)) f(w_z) \, \mathrm{d}w_z, \tag{5}$$

where $\xi'$ is given by

$$\xi'(\sigma, \pi) = \left(1 + \frac{2\beta_{\mathrm{gal}}}{3} + \frac{\beta_{\mathrm{gal}}^2}{5}\right) \xi_0(r) P_0(\mu)$$

$$+ \left(4\frac{\beta_{\mathrm{gal}}}{3} + \frac{4\beta_{\mathrm{gal}}^2}{7}\right) \xi_2(r) P_2(\mu)$$

$$+ \frac{8\beta_{\mathrm{gal}}^2}{35} \xi_4(r) P_4(\mu), \tag{6}$$

where $P_l(\mu)$ are Legendre polynomials, $\mu = \cos(\theta)$ and $\theta$ is the angle between $r$ and $\pi$. $\xi_0(r)$, $\xi_2(r)$ and $\xi_4(r)$ are the monopole, quadrupole and hexadecapole components of the linear $\xi(r)$. In general they are given by (Matsubara & Suto 1996)

$$\xi_{2l}(r) = \frac{-1^l}{r^{2l+1}} \left(\int_0^r x \, \mathrm{d}x\right)^l \left(\frac{\mathrm{d}}{\mathrm{d}x} \frac{1}{x}\right)^l x \xi(x). \tag{7}$$

The effect of RSDs is affirmed when fitting this RSD model, as shown by the lower panels of Fig. 10. The fitting is performed by applying the RSD model to the power-law fit given in Fig. 7(b) (i.e. $r_0 = 2.41 \, h^{-1}$ Mpc and $\gamma = 1.52$). We fit the model first to the real-space $\xi(\sigma, \pi)$ in order to constrain any geometric effects on the 2D clustering that may mimic RSD. The model fitting applied in real space gives best-fitting parameters of $\langle w_z^2 \rangle^{1/2} = 0_{-0}^{+30} \, \mathrm{km \, s^{-1}}$ and $\beta_{\mathrm{gal}} = 0.00_{-0.00}^{+0.06}$, consistent with this measurement having been made in real space. Performing the same fitting to the redshift-space result returns best-fitting values of $\langle w_z^2 \rangle^{1/2} = 160_{-35}^{+45} \, \mathrm{km \, s^{-1}}$ and $\beta_{\mathrm{gal}} = 0.47 \pm 0.22$. From the measured bias for the galaxy sample of $b = 1.85$, we predicted an infall parameter value for this galaxy sample of $\beta_{\mathrm{gal}} = 0.53 \pm 0.03$. Additionally, from the ratio of $\xi(s)/\xi(r)$, we find $\beta_{\mathrm{gal}} = 0.35$, which again is within the $1\sigma$ errors of the 2D fitting result. As for the velocity dispersion, we find that the result is $>1\sigma$ higher than the result for the 1D clustering measurement ($\langle w_z^2 \rangle^{1/2} = 104 \, \mathrm{km \, s^{-1}}$), but is consistent with the intrinsic velocity dispersion measured from the galaxy sample directly ($\langle w_z^2 \rangle^{1/2} = 172 \, \mathrm{km \, s^{-1}}$).

Turning to the $M_* > 10^9 \, h^{-1} \, \mathrm{M_\odot}$ galaxy sample, the top panels of Fig. 9 show $\xi(\sigma, \pi)$ in real (left panel) and redshift (right panel) space (with the lower panels showing the error contours). The $\chi^2$ contours for the fits to the real- and redshift-space measurements are shown in the top panels of Fig. 10. The fitting was again made based on the $\xi(r)$ power-law fit (i.e. $r_0 = 4.16 \, h^{-1}$ Mpc and $\gamma = 1.56$). The best fit for real space is $\beta_{\mathrm{gal}} = 0.00_{-0.00}^{+0.04}$ and velocity dispersion $\langle w_z^2 \rangle^{1/2} = 0_{-0}^{+60} \, \mathrm{km \, s^{-1}}$ with reduced $\chi^2 = 0.7$. In redshift space, we

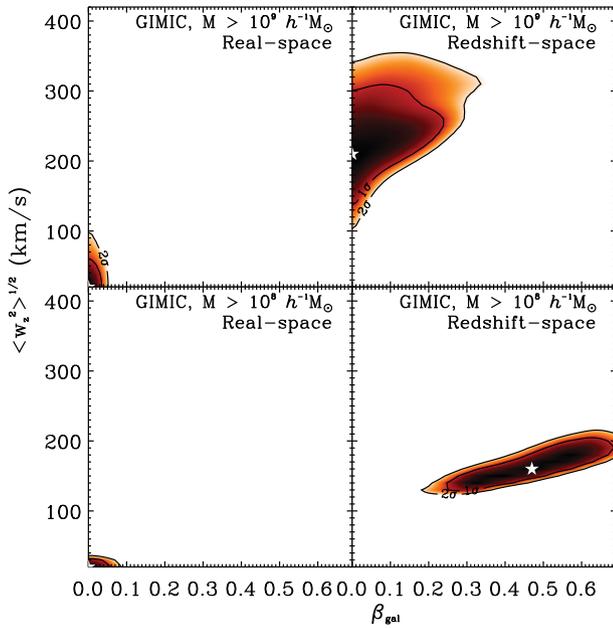**Figure 10.** The top panels show the RSD fitting results in real (left) and redshift space (right) for the GIMIC $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ galaxy sample. The real-space fitting is consistent with the lack of velocity effects in the data, giving best-fitting parameters of $\langle w_z^2 \rangle^{1/2} = 0^{+60}_{-0} \, \mathrm{km \, s^{-1}}$ and $\beta_{\mathrm{gal}} = 0.00^{+0.04}_{-0.00}$. In the redshift-space measurement, we find a velocity dispersion of $\langle w_z^2 \rangle^{1/2} = 210^{+90}_{-70} \, \mathrm{km \, s^{-1}}$. The large-scale motion is constrained as $\beta_{\mathrm{gal}} = 0.00^{+0.24}_{-0.00}$. The lower panels show the best-fitting parameters to the real- and redshift-space results using the RSD model described in the text for the GIMIC $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxy sample. Fitting to the real-space result gives parameters consistent with the null velocity field, with $\langle w_z^2 \rangle^{1/2} = 0^{+30}_{-0} \, \mathrm{km \, s^{-1}}$ and $\beta_{\mathrm{gal}} = 0.00^{+0.06}_{-0.00}$ (left panel). Applying the same model to the redshift-space $\xi(\sigma, \pi)$ we, retrieve best-fitting parameters of $\langle w_z^2 \rangle^{1/2} = 160^{+45}_{-35} \, \mathrm{km \, s^{-1}}$ and $\beta_{\mathrm{gal}} = 0.47 \pm 0.22$ (right panel), consistent with the simulated velocity field.

found $\beta_{\mathrm{gal}} = 0.00^{+0.24}_{-0.00}$ and $\langle w_z^2 \rangle^{1/2} = 210^{+90}_{-70} \, \mathrm{km \, s^{-1}}$ with reduced $\chi^2 = 0.7$.

The bias of $b = 2.80$ suggests a value of $\beta_{\mathrm{gal}} \approx \Omega_m^{0.6}/b = 0.35 \pm 0.02$, which is $>1\sigma$ different from the best-fitting parameter given by the $\xi(\sigma, \pi)$ fitting. The fitted value of $\beta_{\mathrm{gal}} = 0.00^{+0.24}_{-0.00}$ is, however, consistent at the $\approx 1\sigma$ level with the $\beta_{\mathrm{gal}} = 0.24$ implied by the ratio of $\xi(s)/\xi(r)$. In terms of the velocity dispersion fitting parameters, the 1D and 2D fitted $\langle w_z^2 \rangle^{1/2}$ values ($\langle w_z^2 \rangle^{1/2} = 142 \, \mathrm{km \, s^{-1}}$ and $\langle w_z^2 \rangle^{1/2} = 210^{+90}_{-70} \, \mathrm{km \, s^{-1}}$ respectively) are consistent at $\sim 1\sigma$, although the 2D result is again higher than the $1\sigma$ result. The fitting results are summarized in Table 3.

In summary, the analysis of $\xi(\sigma, \pi)$ from the simulation has shown that we may determine RSD effects using the 2D clustering consistently (at the $\sim 1\sigma$ level) with the analysis of the 1D clustering. There is some tension for the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample, where the best-fitting $\beta_{\mathrm{gal}}$ is zero, however this is still consistent with

**Table 3.** Results for the power-law fits to the 2D galaxy auto-correlation functions.

| Sample | $\beta_{\mathrm{gal}}$ | $\langle w_z^2 \rangle^{1/2}$ (km s$^{-1}$) |
|---|---|---|
| VLRS (Bielby et al. 2013) | $0.38 \pm 0.19$ | $420^{+140}_{-160}$ |
| GIMIC $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ | $0.47 \pm 0.22$ | $160^{+45}_{-35}$ |
| GIMIC $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ | $0.00^{+0.24}_{-0.00}$ | $210^{+90}_{-70}$ |

the 1D clustering analysis at the $1\sigma$ level. In all cases, the model successfully constrains the real-space clustering to be consistent with there being no RSD effects. In addition, the infall-parameter results are consistent with the linear theory analysis at the $1\sigma$ level in the case of the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ sample and the $2\sigma$ level for the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample.

Further to this, we have shown that the GIMIC galaxy population has properties consistent with observations of LBGs at $z \sim 3$. For example, Bielby et al. (2013) presented the results for $\xi(\sigma, \pi)$ for $z \sim 3$ LBGs, finding $\beta(z = 3) = 0.38 \pm 0.19$, with $r_0 = 3.83 \pm 0.24 \, h^{-1}$ Mpc and $\gamma = 1.60 \pm 0.09$. The $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ galaxy clustering gives consistent values for all three of these parameters at the $1\sigma$ level. Unfortunately, the small-scale velocity field for the observations is dominated by redshift errors, rather than the intrinsic galaxy peculiar velocities, so we have no suitable $z \sim 3$ data to compare our small-scale results with. However, the results obtained from the simulation for $\langle w_z^2 \rangle^{1/2}$ are instructive for observational analyses.

## 5 GALAXIES AND THE IGM

As discussed earlier, the relationship between the galaxy population and the IGM is key to understanding galaxy growth and evolution. Galaxies require large haloes of gas in order to grow to the large masses we observe at the present day, whilst the supply and regulation of the flow of gas into galaxies dictates the distribution of galaxy masses we observe.

From observations of galaxy winds with speeds of $\gtrsim 300 \, \mathrm{km \, s^{-1}}$ for the LBG population (e.g. via the offset nebulae and interstellar medium spectral features), it is evident that outflowing material exists in these star-forming galaxies (e.g. Pettini et al. 2001; Shapley et al. 2003; Bielby et al. 2011). A number of authors have thus attempted to detect the effects of such outflows on the distribution of gas around the $z \sim 2$–3 star-forming galaxy population via the Ly$\alpha$ forest observed in the spectra of background sightlines (e.g. A03; A05; Crighton et al. 2011; Rudie et al. 2012; Rakic et al. 2012).

In this section, we perform an analysis of the cross-correlation between galaxies and the Ly$\alpha$ forest using both the VLRS observational data and the GIMIC simulation. We apply the same dynamical models as in the previous sections to the cross-correlation analysis. In the case of the galaxy–Ly$\alpha$ cross-correlation, the relation between redshift- and real-space correlations will become (Mountrichas, Sawangwit & Shanks 2009)

$$\xi(s)/\xi(r) = \left[ 1 + \frac{1}{3}(\beta_{\mathrm{gal}} + \beta_{\mathrm{Ly}\alpha}) + \frac{1}{5}\beta_{\mathrm{gal}}\beta_{\mathrm{Ly}\alpha} \right]. \quad (8)$$

The linear bias of the gas obtained from $b^2 = \xi_{\mathrm{Ly}\alpha}/\xi_{\mathrm{DM}}$ is $b \approx 0.3$ (see Section 6), but this is not the bias required to assess the effect of gas infall via $\beta_{\mathrm{Ly}\alpha}$. This is because of the non-linear relation $F = e^{-\tau}$ between Ly$\alpha$ transmission and optical depth, $\tau$, where most of the physics in the Ly$\alpha$ forest is contained in $\tau$. According to Mc-Donald et al. (2000, 2003), the infall parameter $\beta_{\mathrm{Ly}\alpha} = \Omega_m^{0.6} \times b_\eta/b_\delta$ and $b_\eta$ and $b_\delta$ have to be determined from simulations. McDonald (2003) found results for $\beta_{\mathrm{Ly}\alpha} = 1$–1.6, depending on the resolution of the simulations. We therefore take $\beta_{\mathrm{Ly}\alpha} = 1.3$ as our estimate of the gas dynamical infall parameter. McDonald (2003) did not use the RSD techniques used here, so this and the fact that we are using a higher resolution SPH simulation make it interesting to check whether linear theory with their $\beta_{\mathrm{Ly}\alpha}$ fits our simulated data. McDonald et al. (2000) argue that the form of the flux correlation function is proportional to the mass correlation function in the

linear regime. Following McDonald (2003), we shall assume that we can take account of 'finger-of-God' velocity dispersions in the usual way by convolving the transmission correlation function with a Gaussian of the appropriate dispersion.

We perform the LBG–Lyα cross-correlation using the normalized pixel flux values along the quasar sightlines, where the normalized flux or transmissivity is given by

$$T = \frac{\bar{T}(z=3)}{\bar{T}(z)} \frac{f}{f_{con}}, \tag{9}$$

where $f$ is the observed flux at a given wavelength/Lyα redshift and $f_{con}$ is the flux continuum at that wavelength/Lyα redshift. Following A03, our derived values of $T$ incorporate a renormalization to remove the redshift evolution from the normalized flux based on $\bar{T}$, which is given by

$$\bar{T}(z) = 0.676 - 0.220(z - 3), \tag{10}$$

where $z$ is the redshift of a given pixel (McDonald et al. 2000). We do not include the forest at wavelengths below the intrinsic Lyβ emission of the quasars, in order to avoid regions contaminated by Lyβ absorption lines. Thus, only the spectrum between the Lyβ and Lyα is used in this calculation. We also excluded the wavelength range within 20 Å of the intrinsic Lyα emission to avoid any proximity effects from the quasars.

We then use the transmissivity of the Lyα forest as calculated above to perform the LBG–Lyα cross-correlation function. The LBG–Lyα cross-correlation function is calculated from

$$\langle T(s) \rangle = \frac{\langle DT(s) \rangle}{N(s)}, \tag{11}$$

where $\langle DT(s) \rangle$ is the number of galaxy–Lyα pairs weighted by the normalized transmissivity for each separation. $N(s)$ is the number of LBGs that contribute to the cross-correlation function at each separation.

## 5.1 Observed LBG–Lyα cross-correlation

### 5.1.1 1D cross-correlation, $\langle T(s) \rangle$

In Fig. 11, we present the latest result for the LBG–Lyα cross-correlation from the VLRS (left panel: asterisks). This covers a

broad range of scales, measuring to separations of $s \approx 20\,h^{-1}$ Mpc. Errors on the data points are calculated by taking the standard deviation of the $\langle T(s) \rangle$ measure across all the individual galaxies contributing to a given bin, divided by the square root of the number of galaxies contributing to that bin. We see an overall continuous decrease in Lyα transmission down to the minimum scale probed of $s = 0.25\,h^{-1}$ Mpc (although this smallest bin contains only a single galaxy).

We also show the LBG–Lyα transmissivity correlation function for the publicly available Keck data that we incorporate into our 2D analysis (centre panel: pink diamonds) and the A05 result combined with our own VLRS result (right panel: blue circles). In each panel, we also show the results of A03 (grey triangles) and A05 (grey squares). We note in passing that our own reductions of the Keck sample HIRES data give results consistent with the A03 LBG–Lyα results. At separations below $s \approx 5\,h^{-1}$ Mpc, the combined sample has the same trend as A05, with no evidence for a turn-up at $s < 1\,h^{-1}$ Mpc, a feature that was claimed by A03 to be evidence for feedback. With the larger sample of LBGs close to quasar sightlines compared with Crighton et al. (2011), we have now strengthened the evidence against feedback strongly decreasing Lyα absorption on $s \lesssim 1\,h^{-1}$ Mpc scales around galaxies.

### 5.1.2 2D cross-correlation, $\xi(\sigma, \pi)$

We now use the latest VLRS data sample of $\approx 2000$ LBGs alongside the Keck-based LBG–Lyα data set to measure the 2D LBG–Lyα cross-correlation, $\xi(\sigma, \pi)$. By combining these two surveys, we can compare the correlation functions for a wider range of separations than would otherwise be possible (the VLRS giving 2–3 times the coverage on the $\sigma$ scale compared with the Keck data). The LBG–Lyα $\xi(\sigma, \pi)$ from the Keck+VLRS sample is presented in Fig. 12.

In order to fit the RSD model to these data, we first need an estimate of the real-space autocorrelation function. The double power-law fit to $\langle T(s) \rangle$ is unsuitable, as it contains within it the imprint of the RSD effects. We therefore follow the usual route to estimating the real-space clustering and calculate the projected correlation function, $w_p(\sigma)$. This is calculated by integrating the 2D correlation function along the line-of-sight direction, $\pi$:

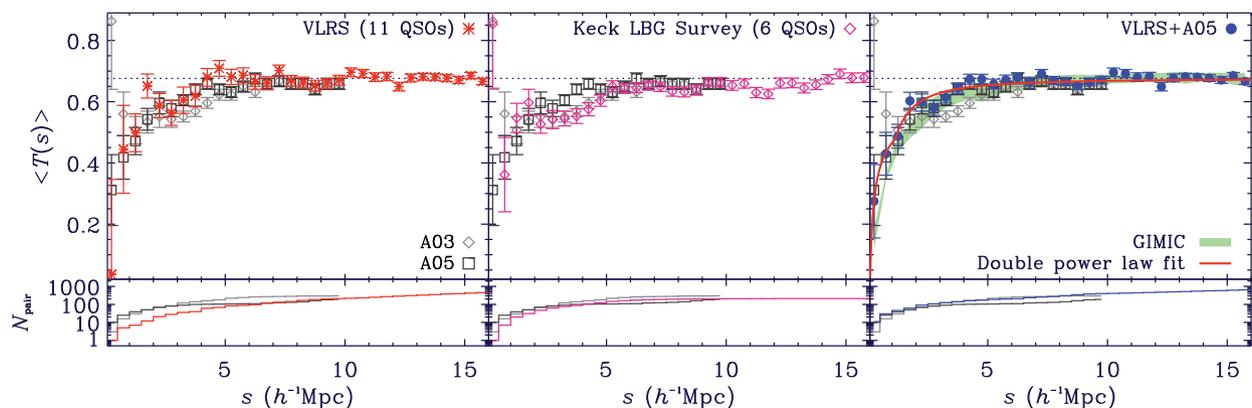$$w_p(\sigma) = 2 \int_0^\infty \xi(\sigma, \pi)\, d\pi. \tag{12}$$



**Figure 11.** The mean Lyα transmissivity as a function of distance, *s*, from galaxies in observed $z \approx 3$ samples. The left-hand panel shows the result for the 11 sightlines observed as part of the VLRS alone, the central panel shows the result for the six sightlines observed with Keck and the right-hand panel shows the result for the VLRS result combined with the result of A05. The lower panels in each case show the number of galaxy–sightline pairs within a given separation. In each of the panels, we also show the results of A03 (grey diamonds) and A05 (grey squares) for comparison. In the right-hand panel, we also show the result of a double power-law fit to the VLRS + Keck data (curve), the parameters for which are given in Table 4. We also show in the right-hand panel the $\langle T(s) \rangle$ result from the GIMIC analysis given by the green shaded region. See the online paper for a colour version of this figure.
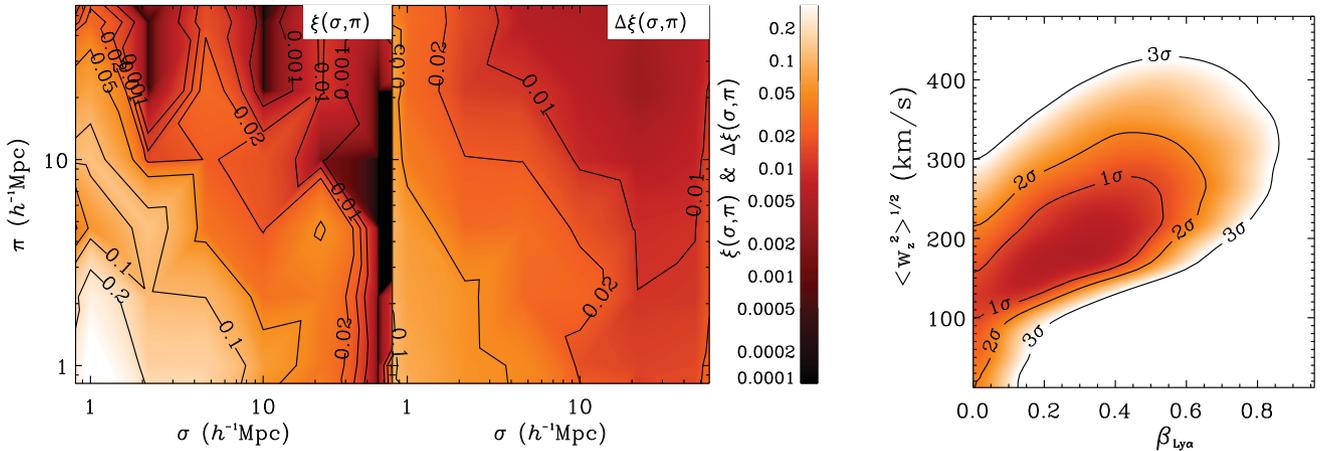
**Figure 12.** The left-hand panels show the LBG–Lyα $\xi(\sigma, \pi)$ and jackknife errors in $\xi(\sigma, \pi)$ for the combined Keck+VLRS data. The right-hand panel shows the result of fitting the $\xi(\sigma, \pi)$ model to the data, with best-fitting parameters given by $\beta_{Ly\alpha} = 0.33^{+0.23}_{-0.33}$ and $\langle w_z^2 \rangle^{1/2} = 190 \pm 90\,\mathrm{km\,s^{-1}}$ (assuming an underlying double power-law form as given in Table 4 and with $\beta_{gal} = 0.38$).
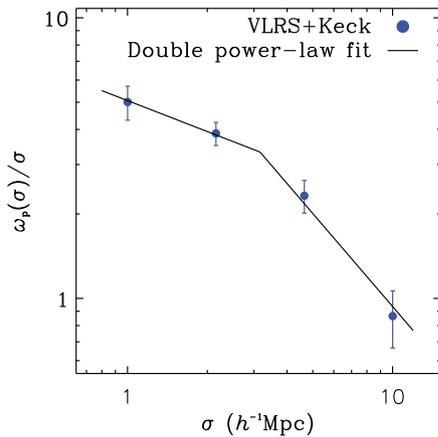


**Figure 13.** The observed LBG–Lyα $w_\mathrm{p}(\sigma)/\sigma$ result based on the VLRS+Keck galaxy and quasar sightline data set (filled blue cirlces). A double power-law fit is shown, the parameters for which are given in Table 4.

The result is shown in Fig. 13. At small scales ($\sigma \lesssim 2\,h^{-1}$ Mpc), the clustering measurement will have a flatter slope due to the satu-ration of Lyα lines in the forest and so we fit the $w_\mathrm{p}(\sigma)$ measurement with a double power law. This is only marginally necessary given the error estimates for the measured $w_\mathrm{p}(\sigma)$ data points and is in part motivated by the analysis of the simulated sightlines that follows. Each power law takes the form

$$\frac{w_\mathrm{p}(\sigma)}{\sigma} = C\xi(\sigma) = C\left(\frac{r_0}{\sigma}\right)^\gamma, \qquad (13)$$

where $C$ is given by

$$C = \frac{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{\gamma-1}{2}\right)}{\Gamma\left(\frac{\gamma}{2}\right)}. \qquad (14)$$

The resulting best-fitting parameters assuming this double power law are given in Table 4. We then use this fit as the basis with which to fit for RSD in the $\xi(\sigma, \pi)$ measurement. As in the galaxy–galaxy autocorrelation analysis, we use a model incorporating a Gaussian form for the effects of pairwise velocities, characterized by $\langle w_z^2 \rangle^{1/2}$, but now with the large-scale infall characterized by a combination of $\beta_{Ly\alpha}$ and $\beta_{gal}$ (where $\beta_{gal}$ is constrained by the autocorrelation results). The model is identical to that described earlier, except that equation (6) is now replaced by

$$\xi'(\sigma, \pi) = \left(1 + \frac{\beta_{gal} + \beta_{Ly\alpha}}{3} + \frac{\beta_{gal}\beta_{Ly\alpha}}{5}\right)\xi_0(r)P_0(\mu)$$
$$+ \left(2\frac{\beta_{gal} + \beta_{Ly\alpha}}{3} + \frac{4\beta_{gal}\beta_{Ly\alpha}}{7}\right)\xi_2(r)P_2(\mu)$$
$$+ \frac{8\beta_{gal}\beta_{Ly\alpha}}{35}\xi_4(r)P_4(\mu), \qquad (15)$$

where $P_l(\mu)$ are again Legendre polynomials, $\mu = \cos(\theta)$ and $\theta$ is the angle between $r$ and $\pi$. $\xi_0(r)$, $\xi_2(r)$ and $\xi_4(r)$ are the monopole, quadrupole and hexadecapole components of the linear $\xi(r)$ and are given in equation (7).

The resulting $\Delta\chi^2$ contours for this fit are shown in the right-hand panel of Fig. 12, with the best-fitting result given by $\beta_{Ly\alpha} = 0.33^{+0.23}_{-0.33}$ and $\langle w_z^2 \rangle^{1/2} = 190 \pm 90\,\mathrm{km\,s^{-1}}$ (given $\beta_{gal} = 0.38$).

As discussed, McDonald (2003) predict a value for the infall parameter for the Lyα forest at $z = 3$ of $\beta_{Ly\alpha} = 1.3 \pm 0.3$. Our

**Table 4.** Results for the power-law fits to the 1D galaxy–Lyα cross-correlation functions.

| Sample | $s_{0,\,s}$ ($h^{-1}$ Mpc) | $\gamma_s$ | $s_{0,\,l}$ ($h^{-1}$ Mpc) | $\gamma_l$ | $\beta_{Ly\alpha}$ |
|---|---|---|---|---|---|
| VLRS – from $\xi(s)$ | $0.08 \pm 0.04$ | $0.47 \pm 0.10$ | $0.49 \pm 0.32$ | $1.47 \pm 0.91$ | — |
| Sample | $r_{0,\,s}$ ($h^{-1}$ Mpc) | $\gamma_s$ | $r_{0,\,l}$ ($h^{-1}$ Mpc) | $\gamma_l$ | $\beta_{Ly\alpha}$ |
| VLRS – from $w_\mathrm{p}(\sigma)$ | $0.020^{+0.074}_{-0.018}$ | $0.37^{+0.45}_{-0.14}$ | $0.59^{+0.90}_{-0.20}$ | $1.10 \pm 0.74$ | — |
| GIMIC $M_* > 10^8\,h^{-1}\,\mathrm{M_\odot}$ | $0.10 \pm 0.07$ | $0.46 \pm 0.22$ | $0.51 \pm 0.39$ | $1.25 \pm 0.61$ | $0.27 \pm 0.05$ |
| GIMIC $M_* > 10^9\,h^{-1}\,\mathrm{M_\odot}$ | $0.16 \pm 0.09$ | $0.46 \pm 0.19$ | $0.61 \pm 0.34$ | $1.18 \pm 0.43$ | $0.31 \pm 0.07$ |

measured value of $\beta_{Ly\alpha} = 0.33$ is more than $3\sigma$ lower than this predicted value. Comparing this with other observations, Slosar et al. (2011) report a range of $0.44 < \beta_{Ly\alpha} < 1.20$, at central redshift $z = 2.25$, from the analysis of Baryon Oscillation Spectroscopic Survey (BOSS) quasar spectra. This is consistent at the $1\sigma$ level with our result, although at a lower redshift.

Predicting the velocity dispersion, we take the pairwise velocity dispersion measured for the galaxies ($\langle w_z^2 \rangle^{1/2} = 420$ km s$^{-1}$: Bielby et al. 2013), which includes both the intrinsic dispersion and the velocity measurement errors, and combine this with the predicted velocity dispersion measured from the GIMIC simulation earlier (120 km s$^{-1}$). As the galaxy measurement is a 'pairwise' velocity, we thus need to divide this by $\sqrt{2}$ and therefore would expect $\langle w_z^2 \rangle^{1/2} = \sqrt{297^2 + 120^2} = 320$ km s$^{-1}$ for the galaxy–Ly$\alpha$ $\langle w_z^2 \rangle^{1/2}$. The result obtained from the LBG–Ly$\alpha$ $\xi(\sigma, \pi)$ is consistent with this predicted value within the $2\sigma$ contours. We shall return to these Keck+VLRS results for comparison with the results from the GIMIC simulations described below.

This measurement of the 2D LBG–Ly$\alpha$ cross-correlation is one of only a few such measurements and the only one to give a full parametrized model fitting to the RSD. Rakic et al. (2012) and Turner et al. (2014) show the 2D LBG–H I pixel–optical-depth (POD) cross-correlation, giving estimated velocity dispersions of $\langle w_z^2 \rangle^{1/2} \sim 240$ and $\sim 260$ km s$^{-1}$ respectively. There are significant differences between our analysis and the work of these two authors, not least that they analyse a broader range in optical depth by including higher order Lyman series lines, but we note that our measured velocity dispersion is consistent with their results.

## 5.2 LBG–Ly$\alpha$ cross-correlation from simulations

As with the data, we compute the LBG–Ly$\alpha$ cross-correlation using the methods described above. We note, however, that the renormalization to $z = 3$, given by equation (10), is redundant here given that the simulated gas and galaxies are all at the same epoch already.

### 5.2.1 Coherent motion of gas and galaxies

In the top panels of Fig. 14, we show the Ly$\alpha$ mean transmissivity as a function of sightline–galaxy separation for the $M_* > 10^8\,h^{-1}\,M_\odot$ (left panel) and $M_* > 10^9\,h^{-1}\,M_\odot$ (right panel) galaxy samples and for three combinations of the gas and galaxies from the GIMIC simulation: galaxies in real space with Ly$\alpha$ in redshift space ($r - z$, green diamonds); galaxies in redshift space with Ly$\alpha$ in real space ($z - r$, yellow triangles); and both the galaxies and Ly$\alpha$ in redshift space ($z - z$, squares). As in previous plots of $\langle T(s) \rangle$, the results are scaled to the mean transmissivity at $z = 3$ (i.e. $\bar{T}(z = 3) = 0.676$). It is interesting to note that the decrease to smaller scales is enhanced as we go from the $r - z$ (or $z - r$) combination to the $z - z$ combination. If we assume that random Gaussian motions dominate galaxy peculiar motions, then this is a surprising result. The same effect is seen for both the $M_* > 10^8\,h^{-1}\,M_\odot$ and $M_* > 10^9\,h^{-1}\,M_\odot$ galaxy samples. This is, however, simply the result of a large ($\sim$100 km s$^{-1}$) bulk flow of material within the simulation volume, i.e. the analysis has not been performed in the mean rest frame of the particles in the box. Indeed, this bulk motion is clearly evident in Fig. 2.

The lower panels of Fig. 14 again show the galaxy–Ly$\alpha$ mean transmissivity as a function of sightline–galaxy separation for the
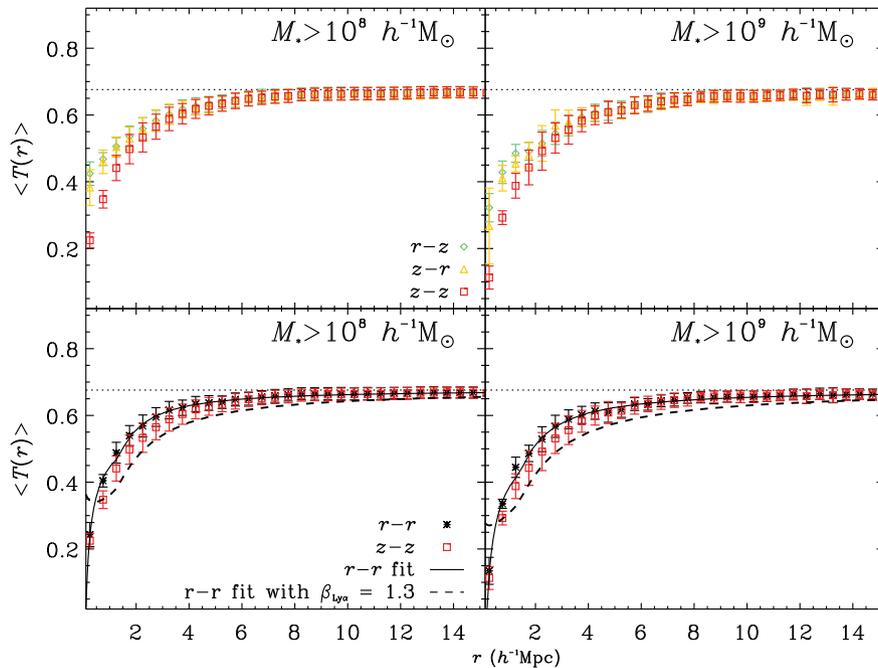


**Figure 14.** The transmissivity profile, $\langle T(r) \rangle$, around simulated galaxies within GIMIC for the $M_* > 10^8\,h^{-1}\,M_\odot$ (left-hand panels) and $M_* > 10^9\,h^{-1}\,M_\odot$ (right-hand panels) galaxy samples. The top panels show the cross-correlation calculated using combinations of the galaxies in real space with the gas in redshift space (diamonds), the galaxies in redshift space with the gas in real space (triangles) and both galaxies and gas in redshift space (squares). The lower panels show the same redshift-space galaxy–Ly$\alpha$ cross-correlation (squares) compared with the real-space cross-correlation (black asterisks). For both the $M_* > 10^8\,h^{-1}\,M_\odot$ (left) and $M_* > 10^9\,h^{-1}\,M_\odot$ (right) results, we show a double power-law fit to the real-space result and this fit convolved with the predicted RSDs (i.e. $\beta_{Ly\alpha} = 1.3$ with $\langle w_z^2 \rangle^{1/2} = 139$ km s$^{-1}$ and $\beta_{gal} = 0.53$ for the $M_* > 10^8\,h^{-1}\,M_\odot$ sample; $\langle w_z^2 \rangle^{1/2} = 156$ km s$^{-1}$ and $\beta_{gal} = 0.35$ for the $M_* > 10^9\,h^{-1}\,M_\odot$ sample). All error bars were calculated using the jackknife method.

$M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ (left panel) and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ (right panel) galaxy samples, but this time for combinations of both galaxies and Ly$\alpha$ in real space ($r - r$, blue asterisks) and both galaxies and Ly$\alpha$ in redshift space ($z - z$, red squares).

Focusing on the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxies, we see that both the $r - r$ and $z - z$ results show the same trends. At a distance $r > 5 \, h^{-1}$ Mpc, the measured $\langle T(r) \rangle$ increases towards the mean value. As separations decrease below $5 \, h^{-1}$ Mpc, the transmissivity decreases, indicating an increase in the H I density as we approach the galaxy. In terms of the effects of RSD, at separations of $r \sim 1$–$6 \, h^{-1}$ Mpc we see that the galaxy–Ly$\alpha$ transmissivity correlation function in redshift space lies lower than the real-space cross-correlation function. This behaviour is suggestive of the impact of coherent infall on the measured $z - z$ cross-correlation function.

To investigate this further, we perform a fit to the correlation function using a power-law form given by $\langle T(r) \rangle = (1 - (r_0/r)^\gamma) \, \bar{T}(z = 3)$. We first attempted a single power-law fit, but found that this failed to match both the large- and small-scale trends. This is primarily due to the non-linear nature of the relationship between the normalized flux measurement and the gas density, whereby, given a high enough column density of neutral hydrogen, the absorption line will reach zero flux and saturate. At this point, the normalized flux no longer gives a measure of increasing gas density and simply asymptotes to a value of zero. This has a significant effect on our measure of $\langle T(r) \rangle$, where close to galaxies the increasing mean gas density leads our measure of $\langle T(r) \rangle$ to turn over. We approximate this behaviour with a double power-law function: one power law fitted to the large-scale trend (i.e. $r \gtrsim 1.6 \, h^{-1}$ Mpc) and another to approximate the small-scale curtailing of $\langle T(r) \rangle$. For the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ sample, we find best-fitting parameters of $r_{0,\mathrm{s}} = 0.10 \pm 0.07 \, h^{-1}$ Mpc, $\gamma_\mathrm{s} = 0.46 \pm 0.22$, $r_{0,1} = 0.51 \pm 0.39 \, h^{-1}$ Mpc and $\gamma_1 = 1.25 \pm 0.61$ (where subscript 's' denotes the small-scale power law and subscript 'l' denotes the large-scale power-law parameters). This fit is plotted as the solid black curve in the lower left panel of Fig. 14 (and is summarized in Table 4).

As a first step in analysing the RSD effects on the cross-correlation, we transform this fitted real-space fit to the GIMIC $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ sample cross-correlation with our RSD model and some reasonable estimates of what we may expect the RSD parameters to be. For the galaxy coherent large-scale motion, we have a value of $\beta_\mathrm{gal} = 0.53$ derived from the galaxy–galaxy autocorrelation. For the Ly$\alpha$ coherent large-scale motion, we take $\beta_\mathrm{Ly\alpha} = 1.3$ as predicted by the simulations of McDonald (2003). Finally, for the velocity dispersion parameter, we take $\langle w_z^2 \rangle^{1/2} = \sqrt{(104/\sqrt{2})^2 + 120^2} = 139 \, \mathrm{km \, s^{-1}}$, i.e. combining the measured galaxy velocity dispersion (from Fig. 3) and Ly$\alpha$ velocity dispersion (from Fig. 6) in quadrature. The result is given by the black dashed line in the lower left panel of Fig. 14.

We perform an identical analysis with the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample, fitting a double power law to the real-space $\langle T(r) \rangle$, finding best-fitting parameters of $r_{0,\mathrm{s}} = 0.16 \pm 0.09 \, h^{-1}$ Mpc, $\gamma_\mathrm{s} = 0.46 \pm 0.19$, $r_{0,1} = 0.61 \pm 0.34 \, h^{-1}$ Mpc and $\gamma_1 = 1.18 \pm 0.43$ (shown by the solid black curve in the lower right panel of Fig. 14). The RSD model based on this double power-law fit is shown by the dashed black line in the lower right panel of Fig. 14 and is based on parameter values of $\beta_\mathrm{gal} = 0.35$, $\beta_\mathrm{Ly\alpha} = 1.3$ and $\langle w_z^2 \rangle^{1/2} = \sqrt{(142/\sqrt{2})^2 + 120^2} = 156 \, \mathrm{km \, s^{-1}}$.

It is evident that the selected parameters do not provide a good fit to the redshift-space results from the GIMIC simulation in either case. For both the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ samples, the model overpredicts the effects of the coherent infall (at

scales of $\gtrsim 2 \, h^{-1}$ Mpc) and the velocity dispersion at smaller scales. We investigate this further in the following sections.

### 5.2.2 Dynamical infall in $\xi(r)$

To better visualize any distortions in the cross-correlation, we calculate the function $\xi(r) = 1 - \langle T(r) \rangle / \bar{T}(z = 3)$. The results for $\xi(r)$ are shown in the top panels of Fig. 15. The points and curves are the same as given in the lower panels of Fig. 14 (except transformed from $\langle T(r) \rangle$ to $\xi(r)$) and again the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ samples are shown in the left- and right-hand panels respectively. The models used for the curves are identical to those given in the previous section, but we now see more clearly why a single power law is unable to provide a good fit to the real-space data points in both the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ cases. It is also clearer in these plots how the $\beta_\mathrm{Ly\alpha}, \langle w_z^2 \rangle^{1/2} = 139 \, \mathrm{km \, s^{-1}} / \langle w_z^2 \rangle^{1/2} = 156 \, \mathrm{km \, s^{-1}}$ RSD models provide a poor fit to the redshift-space $\xi(s)$ results (red squares). The model lies at $\sim 1\sigma$ above the data at all points above $\sim 1 \, h^{-1}$ Mpc, whilst it also overpredicts the effects of small-scale velocity dispersion. This is the case for both the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ samples.

It is the $\beta_\mathrm{Ly\alpha} = 1.3$ value that is proving too high here, resulting in the model tending to overpredict the galaxy–Ly$\alpha$ cross-correlation function. In the lower panels of Fig. 15, we show the ratio $\xi(s)/\xi(r)$. We measure a weighted average of the ratio over scales of $1 \geq r \geq 12 \, h^{-1}$ Mpc of $\langle \xi(s)/\xi(r) \rangle = 1.29 \pm 0.02$ and $\langle \xi(s)/\xi(r) \rangle = 1.24 \pm 0.03$ for the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ samples respectively. Via equation (3), these values correspond to $\beta_\mathrm{Ly\alpha} = 0.27 \pm 0.05$ ($M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$) and $\beta_\mathrm{Ly\alpha} = 0.31 \pm 0.07$ ($M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$).

From the $\xi(s)$ measurement, we are thus able to place constraints on a measure of the infall of gas towards galaxies via the $\beta_\mathrm{Ly\alpha}$ quantity, consistently obtaining $\beta_\mathrm{Ly\alpha} \sim 0.3$ for both of our galaxy samples. We now move to the 2D cross-correlation function to evaluate whether we obtain consistent results with a 2D analysis.

### 5.2.3 Dynamical Infall in $\xi(\sigma, \pi)$

We now analyse the properties of the 2D cross-correlation function, $\xi(\sigma, \pi)$. This is calculated in the same way as $\xi(r)$, whilst again we estimate errors in the results using the jackknife method. The GIMIC galaxy–Ly$\alpha$ $\xi(\sigma, \pi)$ results are presented in Fig. 16 for the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxy sample and Fig. 17 for the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample. In each case, the top left panel shows $\xi(\sigma, \pi)$ in real space and the top right panel in redshift space, with the lower panels showing the associated error profiles on the same scale. The dashed long-dashed lines show the $r - r$ double power-law fit to the data, which we use as the input model for our RSD model fitting to the $\xi(\sigma, \pi)$ contours, in which we ascertain the best-fitting values for the parameters $\beta_\mathrm{Ly\alpha}$ and $\langle w_z^2 \rangle^{1/2}$.

We fit both the real- and redshift-space $\xi(\sigma, \pi)$ using the same basic model and limited to a maximum separation of $r = 12 \, h^{-1}$ Mpc (minimizing the impact of the limited simulation size on the results). By first fitting to the real-space results, we provide a baseline test of whether the analysis successfully yields $\beta_\mathrm{Ly\alpha} = 0$ and $\langle w_z^2 \rangle^{1/2} = 0 \, \mathrm{km \, s^{-1}}$ for the case of no peculiar velocities. The $\chi^2$ fitting contours for the real-space measurements are shown in the left-hand panels of Fig. 18 (with the top panel showing the fit to the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ measurement
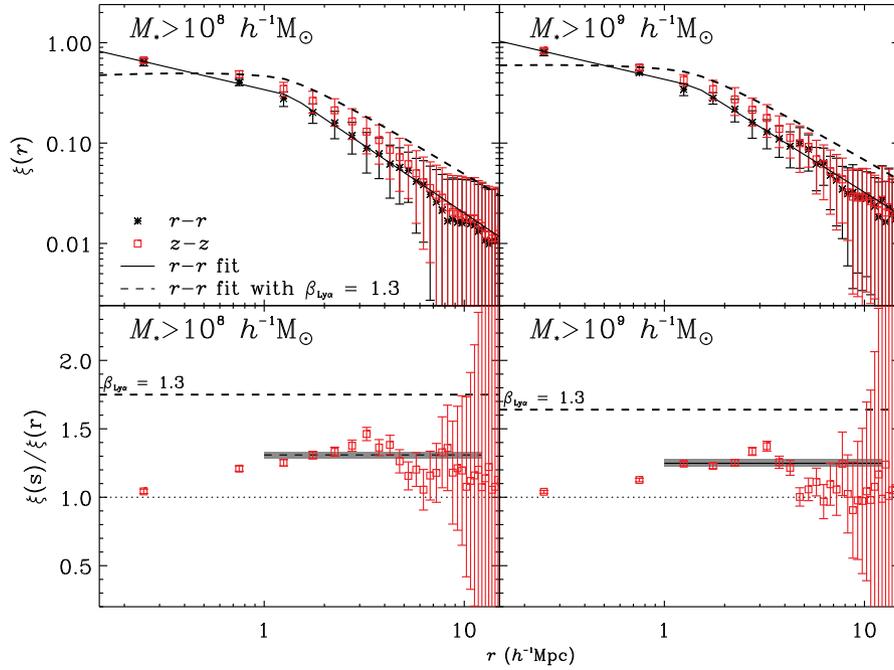
**Figure 15.** The top panels show the LBG–Lyα cross-correlation function, $\xi(r)$, as derived from the $\langle T(r) \rangle$ profiles shown in Fig. 14. Results for the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ sample are shown in the left-hand panels and the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample in the right-hand panels. In each case, we show the same fits as shown in Fig. 14 (solid black curves) and the subsequent redshift-space distorted predictions based on a value of $\beta_{\mathrm{Ly}\alpha} = 1.3$ (dashed black curves). The lower panels show the ratio of the redshift-space cross-correlation functions, $\xi(s)$, to the real-space cross-correlation functions, $\xi(r)$. The dashed black lines in both lower panels shows the large-scale prediction for $\xi(s)/\xi(r)$ assuming $\beta_{\mathrm{Ly}\alpha} = 1.3$. The solid black lines and grey regions show the weighted mean of $\xi(s)/\xi(r)$ measured at $r > 1 \, h^{-1}$ Mpc and the $1\sigma$ errors on the weighted mean.
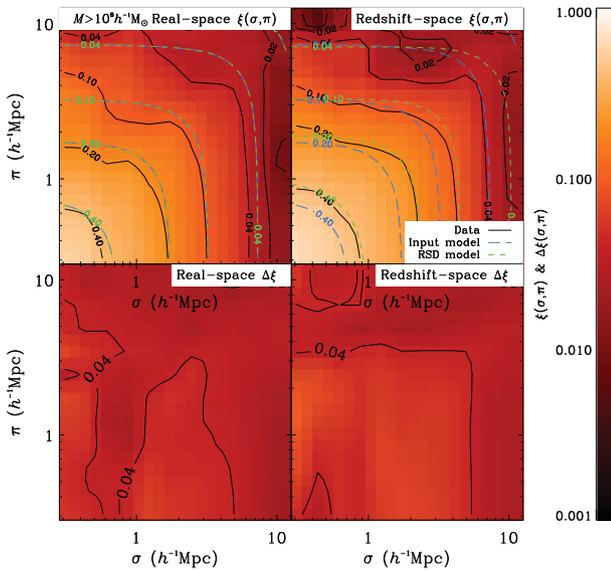


**Figure 16.** The top two panels show the GIMIC galaxy–Lyα $\xi(\sigma, \pi)$ results (shaded map and solid black contours) based on the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ galaxy sample in real space (top-left panel) and in redshift space (top-right panel). We show the underlying double power-law model derived from the real-space correlation function (i.e. with no RSD modelling) by the long-dashed blue contours (identical in the top two panels). The RSD models that best fit the $\xi(\sigma, \pi)$ results (based on this input model) are shown by the short-dashed green contours. Errors were calculated based on a jackknife analysis and are shown in the lower panels.
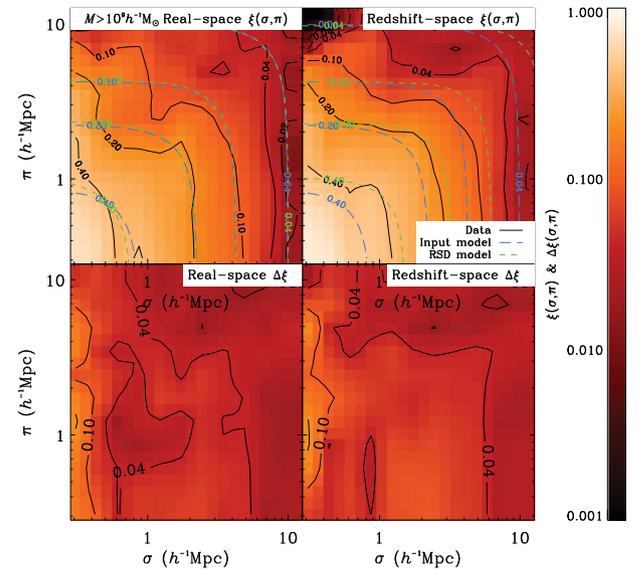
**Figure 17.** As in Fig. 16, but for the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample.

and the lower panel showing the fit to the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ measurement). Starting with the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ real-space result, we find best-fitting parameters entirely consistent with the lack of RSD effects on the $\xi(\sigma, \pi)$ measurement, with $\beta_{\mathrm{Ly}\alpha} = 0^{+0.06}_{-0.00}$ and $\langle w_z^2 \rangle^{1/2} = 0^{+50}_{-0} \, \mathrm{km \, s^{-1}}$. For the $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ measurement, we find $\beta_{\mathrm{Ly}\alpha} = 0^{+0.08}_{-0.00}$ and $\langle w_z^2 \rangle^{1/2} = 50 \pm 25 \, \mathrm{km \, s^{-1}}$. These
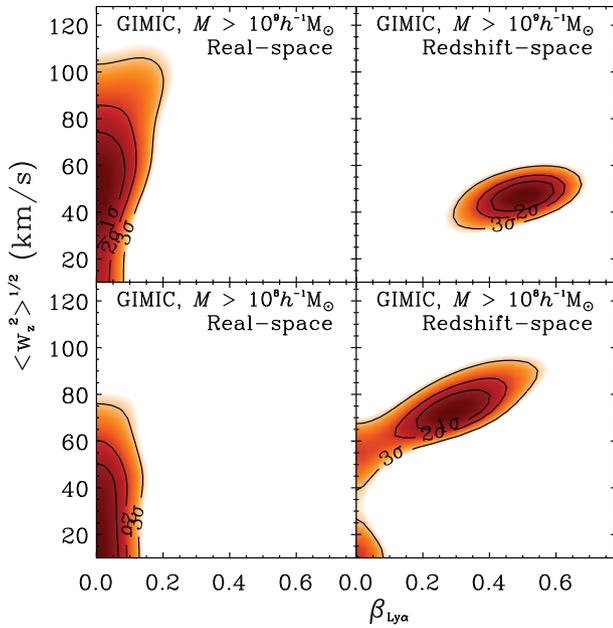
**Figure 18.** The $\Delta\chi^2$ contours for the RSD model fitting to the 2D galaxy–Ly$\alpha$ cross-correlation are shown. The top panels show the fitting results for the $M_* > 10^9\,h^{-1}\,M_\odot$ galaxy sample in real space (left) and redshift space (right). In real space, we find best-fitting parameters of $\beta_{Ly\alpha} = 0.00^{+0.08}_{-0.00}$ and $\langle w_z^2 \rangle^{1/2} = 50 \pm 25\,\mathrm{km\,s^{-1}}$. For the redshift-space $\xi(\sigma, \pi)$, the best-fitting parameters are $\beta_{Ly\alpha} = 0.51 \pm 0.12$ and $\langle w_z^2 \rangle^{1/2} = 48 \pm 7\,\mathrm{km\,s^{-1}}$. The lower panels show the results for the $M_* > 10^8\,h^{-1}\,M_\odot$ GIMIC sample. The best fit for the real-space sample (left panel) is $\beta_{Ly\alpha} = 0.00^{+0.00}_{-0.00}$, $\langle w_z^2 \rangle^{1/2} = 0^{+50}_{-0}\,\mathrm{km\,s^{-1}}$. For the redshift-space $\xi(\sigma, \pi)$, we find $\beta_{Ly\alpha} = 0.28 \pm 0.10$, $\langle w_z^2 \rangle^{1/2} = 73 \pm 9\,\mathrm{km\,s^{-1}}$.

best-fitting models are shown by the short dashed contours in Figs 16 and 17. For the $M_* > 10^8\,h^{-1}\,M_\odot$ sample, the analysis successfully identifies the lack of any velocity information in the result; however, the $M_* > 10^9\,h^{-1}\,M_\odot$ results appears to show a $\sim 2\sigma$ signal for a non-zero $\langle w_z^2 \rangle^{1/2}$, corresponding to some small-scale velocity dispersion. The cause of this is evident from the contour plot of $\xi(\sigma, \pi)$ in the top left panel of Fig. 17, where an extension along the $\pi$ axis is clearly visible at small $\sigma$. This extension is at the $\sim 2\sigma$ level according to the jackknife errors and, given that there are no velocity offsets in this realization, the non-zero result is likely caused by statistical fluctuations at these small $\sigma$ scales. The fact that a zero velocity dispersion is found for the $M_* > 10^8\,h^{-1}\,M_\odot$ sample, in which we have more galaxies, seems to support this conclusion. However, this is important to factor into the analysis when reapplying the model fitting to the redshift-perturbed simulated galaxies.

The $\chi^2$ contours for the model fits to the redshift-space $\xi(\sigma, \pi)$ results are shown in the right-hand panels of Fig. 18, where the top panels show the results for the $M_* > 10^9\,h^{-1}\,M_\odot$ galaxy sample and the lower panels show the results for the $M_* > 10^8\,h^{-1}\,M_\odot$ sample.

For the $M_* > 10^8\,h^{-1}\,M_\odot$ galaxies, we find best-fitting parameters of $\beta_{Ly\alpha} = 0.28 \pm 0.10$ and $\langle w_z^2 \rangle^{1/2} = 73 \pm 9\,\mathrm{km\,s^{-1}}$, whilst for the $M_* > 10^9\,h^{-1}\,M_\odot$ sample we find $\beta_{Ly\alpha} = 0.51 \pm 0.12$ and $\langle w_z^2 \rangle^{1/2} = 48 \pm 7\,\mathrm{km\,s^{-1}}$. The first thing to note is that the measurements for $\beta_{Ly\alpha}$, which should be the same, given they both represent the gas motion, are consistent at the $\approx 1\sigma$ level between the two samples. Further to this, we can perform a comparison with our results from the $\xi(r)$ analysis as a consistency check of the analysis. From $\xi(s)/\xi(r)$, we measured values of $\beta_{Ly\alpha} = 0.27 \pm$

**Table 5.** Results for the power-law fits to the 2D galaxy–Ly$\alpha$ cross-correlation functions.

| Sample | $\beta_{Ly\alpha}$ | $\langle w_z^2 \rangle^{1/2}$ (km s$^{-1}$) |
|---|---|---|
| VLRS+Keck | $0.33^{+0.33}_{-0.23}$ | $190 \pm 90$ |
| GIMIC $M_* > 10^8\,h^{-1}\,M_\odot$ | $0.28 \pm 0.10$ | $73 \pm 9$ |
| GIMIC $M_* > 10^9\,h^{-1}\,M_\odot$ | $0.51 \pm 0.12$ | $48 \pm 7$ |

0.05 and $\beta_{Ly\alpha} = 0.31 \pm 0.07$ from the $M_* > 10^8\,h^{-1}\,M_\odot$ and $M_* > 10^9\,h^{-1}\,M_\odot$ samples respectively. Collating all the measurements of $\beta_{Ly\alpha}$ thus far, the $\beta_{Ly\alpha}$ values are all in strong agreement between the $\xi(r)$ $M_* > 10^8\,h^{-1}\,M_\odot$ and $M_* > 10^9\,h^{-1}\,M_\odot$ results and the $\xi(\sigma, \pi)$ $M_* > 10^8\,h^{-1}\,M_\odot$ result, whilst the $\xi(\sigma, \pi)$ $M_* > 10^9\,h^{-1}\,M_\odot$ result shows some small tension at the $\sim 1.5\sigma$ level.

Now looking to the velocity dispersion results, we find a significant difference between the $M_* > 10^8\,h^{-1}\,M_\odot$ and $M_* > 10^9\,h^{-1}\,M_\odot$ $\xi(\sigma, \pi)$ results. In itself this is not unexpected, given that the galaxy population contributes to this parameter. However, we would expect the $M_* > 10^9\,h^{-1}\,M_\odot$ sample to show a higher velocity dispersion than the $M_* > 10^8\,h^{-1}\,M_\odot$ sample, which is not the case. In addition, the $M_* > 10^9\,h^{-1}\,M_\odot$ redshift-space measurement is itself consistent with the $M_* > 10^9\,h^{-1}\,M_\odot$ real-space measurement, suggesting that we are not actually able to measure the velocity dispersion in this case. Given the small number of pairs at small separations in the $M_* > 10^9\,h^{-1}\,M_\odot$ galaxy–Ly$\alpha$ cross-correlation, this is likely due to the stochastic nature of the signal we are measuring at these small scales. As discussed in Section 5.2.1, based on the measured galaxy and gas velocity distributions we would expect $\langle w_z^2 \rangle^{1/2} \approx 139\,\mathrm{km\,s^{-1}}$ and $\langle w_z^2 \rangle^{1/2} \approx 156\,\mathrm{km\,s^{-1}}$ for the $M_* > 10^8\,h^{-1}\,M_\odot$ and $M_* > 10^9\,h^{-1}\,M_\odot$ samples respectively. This is clearly not the case from our measurements. The explanation may be due to the gas motion being very coherent with the galaxies at small scales. Should the gas and galaxies be moving together in such a way, this could reduce the measured velocity dispersion between the two in the cross-correlation analysis, for $\xi(s)$ as well as $\xi(\sigma, \pi)$. The fitting results are summarized in Table 5.

### 5.2.4 Simulation and observation compared

We next compare the simulated results for galaxy–Ly$\alpha$ $\langle T(s) \rangle$ with the Keck+VLRS data as shown in the right-hand panel of Fig. 11. The GIMIC result for the $M_* > 10^9\,h^{-1}\,M_\odot$ sample in redshift space is shown by the shaded green curve, which follows the observational data points well. The GIMIC result falls to lower values of $\langle T(s) \rangle$ at small scales than the observational result, although only at the $\approx 1\sigma$ level, a potential sign of the effect of observational velocity errors on the data points.

We now compare the simulated results for galaxy–Ly$\alpha$ $\xi(\sigma, \pi)$ with the six Keck quasar + VLRS data as shown in Fig. 12. We have seen that the observed best-fitting parameters for the Keck + VLRS data are $\beta_{Ly\alpha} = 0.33^{+0.33}_{-0.23}$, $\langle w_z^2 \rangle^{1/2} = 190 \pm 90\,\mathrm{km\,s^{-1}}$, with this measurement of the infall parameter being $\approx 3\sigma$ lower than the predicted value of $\beta_{Ly\alpha} = 1.3$. As we have shown above, the best-fitting value for $\beta_{Ly\alpha}$ from the simulated galaxy samples covers a range of $\beta_{Ly\alpha} \approx 0.3$–$0.5$. This simulated $\beta_{Ly\alpha}$ value is consistent within the error estimates of our VLRS observations, but not the theoretically motivated $\beta_{Ly\alpha} = 1.3$. We also note that both the minimum value of $\langle w_z^2 \rangle^{1/2} = 297\,\mathrm{km\,s^{-1}}$ from LBG velocity error and the $\langle w_z^2 \rangle^{1/2} = 320\,\mathrm{km\,s^{-1}}$ value including the full simulated $\langle w_z^2 \rangle^{1/2} = 120\,\mathrm{km\,s^{-1}}$ are consistent with the data at the $\approx 1$–$2\sigma$

level. We conclude that, while the Keck+VLRS $\beta_{Ly\alpha}$ and $\langle w_z^2 \rangle^{1/2}$ estimates are low compared with the initial expectation from theory, they are consistent with the similarly low values of these parameters estimated from $\xi(r)$ and $\xi(\sigma, \pi)$ in the GIMIC simulations.

# 6 AUTO-CORRELATION ANALYSIS OF THE IGM

## 6.1 Lyα 1D autocorrelation function

We now measure the Lyα autocorrelation function in both the observational data and the simulated sightlines, with the aim of again comparing the simulated sightline results with observations and measuring the effect of the velocity field on the clustering. Following Crighton et al. (2011), for each pixel in a quasar line of sight, we calculate

$$\delta = \frac{T}{\bar{T}} - 1, \tag{16}$$

where $T$ and $\bar{T}$ are the measured and the mean normalized flux. We then use this to calculate the autocorrelation function:

$$\xi(\Delta s) = \langle \delta(s)\delta(s + \Delta s) \rangle. \tag{17}$$

For the observational data, we are only able to do this in the line-of-sight direction, as there are only three pairs of quasars that can provide transverse separation measurements and these are all separated by $\gtrsim 20 \, h^{-1}$ Mpc. For the simulated sightlines, we sum all pixels with the separations $\Delta s$, both parallel and perpendicular to the line of sight.

Fig. 19 shows the autocorrelation of Lyα pixels along the line of sight from the observational data. Keck, VLRS and combined samples are presented by diamonds, red asterisks and blue circles, respectively. Error bars were estimated by using the jackknife method. We first compare these with the result from Crighton et al. 2011 (cyan triangles), who measured the autocorrelation us-
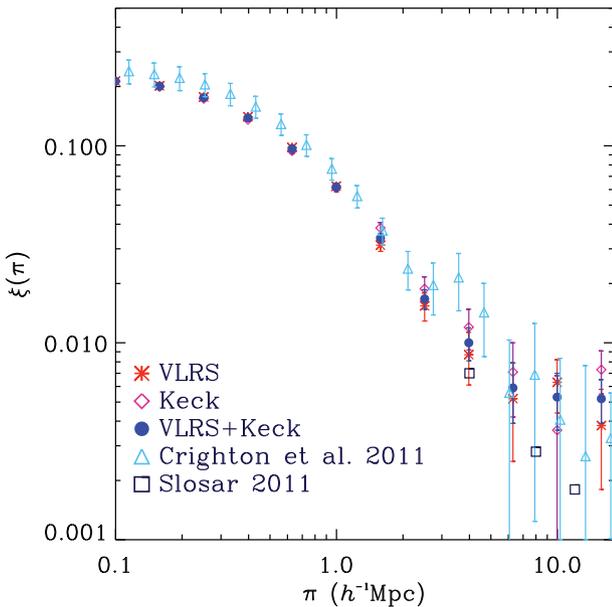
ing seven high-resolution quasars (resolution FWHM $\sim 7 \, \text{km s}^{-1}$). They all show similar results at small scales. We also show the recent BOSS results of Slosar et al. (2011, black squares), which probe scales of $\gtrsim 3 \, h^{-1}$ Mpc and are consistent with the VLRS results.

The autocorrelation functions based on the GIMIC simulated Lyα sightlines are presented in Fig. 20. The real- and redshift-space Lyα autocorrelation functions are shown by black asterisks and squares respectively. The VLRS+Keck result from Fig. 19 is replotted (filled circles) and is found to be consistent with the GIMIC autocorrelation within the quoted error estimates.

Focusing on the simulation, we again see that the redshift- and real-space correlation functions are comparable in amplitude and form. We fit the GIMIC real-space autocorrelation function with a double power-law form as performed with the galaxy–Lyα cross-correlation. The resulting fit is given in Fig. 20 (solid black curve). At small scales, convolving this double power-law fit with a Gaussian of width $120 \times \sqrt{2} = 170 \, \text{km s}^{-1}$, representing the simulation gas peculiar velocity (see Fig. 3), is seen to overestimate the small-scale turnover in the redshift-space correlation function.

Based on the power-law fit at $r > 0.4 \, h^{-1}$ Mpc (and using the relation $b = \sqrt{\xi_{Ly\alpha}/\xi_{DM}}$), the clustering bias of the Lyα forest is $b \approx 0.3$. Assuming $\beta_{Ly\alpha} = \Omega^{0.6}/b$, this bias corresponds to $\beta_{Ly\alpha} \approx 3.3$, which implies $\xi(s)/\xi(r) \approx 5.4$. Again, as noted by McDonald (2003), $\beta_{Ly\alpha}$ has no simple relation to density bias, as for galaxies $\beta_{Ly\alpha}$ has to be estimated from simulations and the simulations of



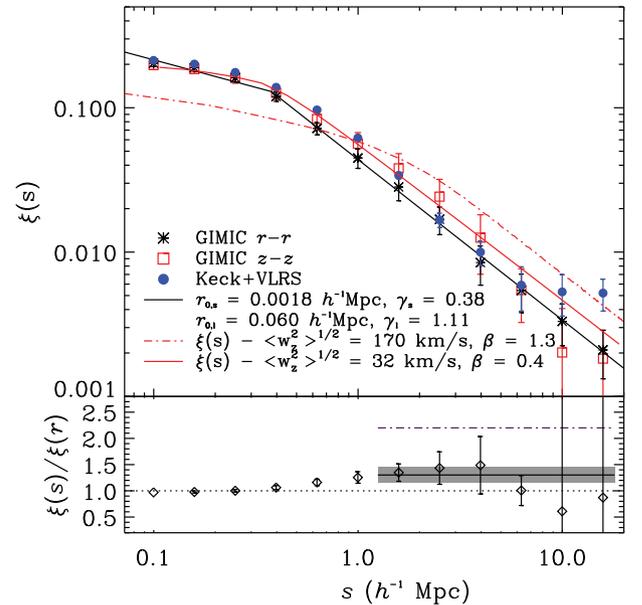**Figure 19.** The autocorrelation of Lyα pixels along the line of sight. The VLRS, Keck and combined samples are shown by red asterisks, pink diamonds and filled circles, respectively. The measurement of Crighton et al. (2011) is also shown (cyan triangles), as is the BOSS result of Slosar et al. (2011, black squares).



**Figure 20.** Top panel: autocorrelation functions of GIMIC Lyα pixels at $z = 3.06$ in the $0\sigma$ simulation. Real-space (black asterisks) and redshift-space (red squares) results are shown. Errors are calculated via the jackknife method. A double power-law fit to the real-space $\xi(r)$ with $r_{0,s} = 0.0018 \pm 0.0015$, $\gamma_s = 0.38 \pm 0.14$, $r_{0,l} = 0.060 \pm 0.034$, $\gamma_l = 1.11 \pm 0.21$ is also shown (black line). The red dot–dashed line is the expected result for the Lyα $\xi(s)$ in redshift space if we convolve the velocity dispersion $\langle w_z^2 \rangle^{1/2} = 170 \, \text{km s}^{-1}$ and $\beta_{Ly\alpha} = 1.3$ with the RSD model. The red solid curve is a model $\xi(s)$ fitted to the GIMIC $z - z$ result and is given by parameter values of $\langle w_z^2 \rangle^{1/2} = 32 \, \text{km s}^{-1}$ and $\beta_{Ly\alpha} = 0.40$ (applied to the double power-law model fit to $\xi(r)$). Bottom panel: GIMIC $\xi(s)/\xi(r)$ with jackknife error bars. The dot–dashed line corresponds to $\xi(s)/\xi(r) = 2.2$ as predicted from linear theory with $\beta_{Ly\alpha} = 1.3$, whilst the solid black line and grey surround show the weighted mean $\xi(s)/\xi(r)$ from the simulation and its $1\sigma$ bounds. See the online paper for a colour version of this figure.

McDonald et al. implied a range $\beta_{\text{Ly}\alpha} = 1$–$1.6$. If we therefore take $\beta_{\text{Ly}\alpha} = 1.3$, then this predicts $\xi(s)/\xi(r) = 2.2$ from equation (3), whereas the simulated value in Fig. 20 is $\xi(s)/\xi(r) = 1.30 \pm 0.14$ (solid black line and shaded region in the lower panel of Fig. 20), which corresponds to $\beta_{\text{gal}} = 0.40 \pm 0.16$. With $\beta_{\text{Ly}\alpha} = 0.40$, the best-fitting velocity dispersion is $\langle w_z^2 \rangle^{1/2} = 32 \pm 7\,\text{km s}^{-1}$. Models where we fixed $\beta_{\text{Ly}\alpha} = 1.3$ and took $\langle w_z^2 \rangle^{1/2} = 170\,\text{km s}^{-1}$ as expected from Fig. 6 are strongly rejected (red dash–dotted curve). With $\beta_{\text{Ly}\alpha} = 1.3$, a best-fitting value of $\langle w_z^2 \rangle^{1/2} = 86^{+11}_{-8}\,\text{km s}^{-1}$ was found, although the model was still rejected in a chi-square test.

Whatever value of $\beta_{\text{Ly}\alpha}$ is chosen, it appears that the fitted value of the velocity dispersion is much lower than we measured in Fig. 6. However, as shown by Crighton et al. (2011), the intrinsic width of the Ly$\alpha$ lines convolved with the instrumental response of the spectrograph can induce artificial autocorrelations at scales $\lesssim 0.7\,h^{-1}$ Mpc, so this effect may contribute to the poor fit of the peculiar velocity RSD model on small scales.

We note that the RSD model for the Ly$\alpha$ auto- and cross-correlation assumes spherical symmetry as we move from real space to redshift space and the Ly$\alpha$ autocorrelation function involves summing along and across quasar lines of sight, which may not be exactly spherically symmetric. However, we shall see that this explanation cannot apply to the Ly$\alpha$ $\xi(\sigma, \pi)$, which we calculate next and which gives consistent results with the $\xi(s)$ analysis.

## 6.2 Ly$\alpha$ 2D autocorrelation function

For each pixel in the Ly$\alpha$ line of sight, we next calculate the Ly$\alpha$ $\xi(\sigma, \pi)$ by using

$$\xi(\sigma, \pi) = \frac{\langle DT(\sigma, \pi) \rangle}{N(\sigma, \pi)}, \tag{18}$$

where $\langle DT(\sigma, \pi) \rangle$ is the number of Ly$\alpha$ pairs weighted by the normalized transmissivity, $T$, for each separation. $N(\sigma, \pi)$ is the number of Ly$\alpha$ pixels that contributed to each pair.

The Ly$\alpha$ $\xi(\sigma, \pi)$ results at $z = 3.06$ for the $0\sigma$ simulation are shown in Fig. 21, with the top left panel showing the result in real space and the top right panel showing the result in redshift space. The associated errors are again shown in the lower panels.
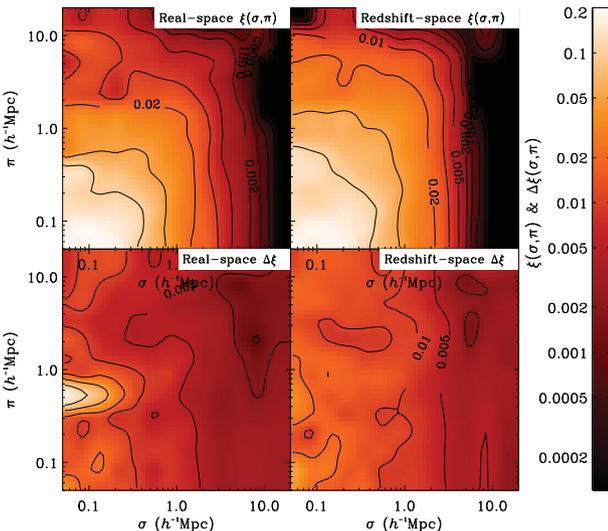


**Figure 21.** The GIMIC Ly$\alpha$ $\xi(\sigma, \pi)$ autocorrelation at $z = 3.06$ in real (top left panel) and redshift space (top right panel). The lower panels show the corresponding jackknife error estimates for the $\xi(\sigma, \pi)$ results.
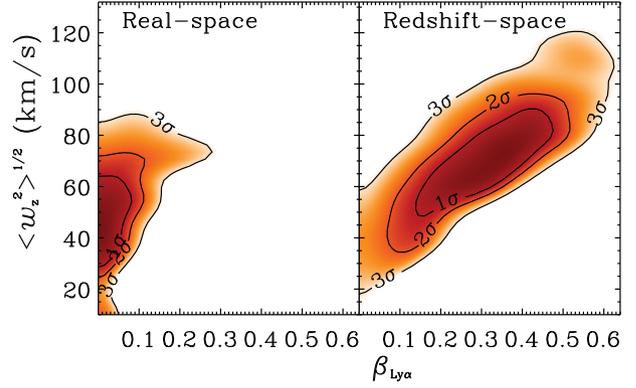
**Figure 22.** Results for the model fits to the GIMIC 2D Ly$\alpha$ autocorrelation functions shown in Fig. 21. The left panel shows the $\Delta\chi^2$ contours for the fit to the real-space $\xi(\sigma, \pi)$, with best-fitting parameters $\beta_{\text{gal}} = 0.00^{+0.09}_{-0.00}$ and $\langle w_z^2 \rangle^{1/2} = 46 \pm 17\,\text{km s}^{-1}$. The right panel shows the same for the redshift-space autocorrelation function. Here the best fit is $\beta_{\text{gal}} = 0.31 \pm 0.17$, $\langle w_z^2 \rangle^{1/2} = 69 \pm 21\,\text{km s}^{-1}$. The fits are based on the underlying double power-law function shown in Fig. 20.

Again we fit the RSD model to the GIMIC results and find $\langle w_z^2 \rangle^{1/2} = 46 \pm 17\,\text{km s}^{-1}$ and $\beta_{\text{gal}} = 0.00^{+0.09}_{-0.00}$ for the real-space result (see left-hand panel of Fig. 22). For redshift space, the best-fitting parameters are the same, with $\beta_{\text{gal}} = 0.31 \pm 0.17$ and $\langle w_z^2 \rangle^{1/2} = 69 \pm 21\,\text{km s}^{-1}$ (see right-hand panel of Fig. 22). We again conclude that the effects of infall in the gas in the GIMIC simulation are much less than predicted from the previous work of McDonald (2003), with an upper limit of $\beta_{\text{Ly}\alpha} \lesssim 0.6$ from $\xi(s)$ and $\beta_{\text{Ly}\alpha} \lesssim 0.5$ from Ly$\alpha$ $\xi(\sigma, \pi)$. Given $\beta_{\text{gal}} = 0.31$, the gas velocity dispersion fit of $\langle w_z^2 \rangle^{1/2} = 69 \pm 21\,\text{km s}^{-1}$ is close to the sub $1h^{-1}$ Mpc value of the velocity dispersion estimated for simulated galaxies, due to correlated motions.

As discussed, comparing the VLRS+Keck $\xi(s)$ result (filled circles) with the GIMIC $\xi(s)$, we found good agreement between the two within the $1\sigma$ errors for the two data sets. However, we do not calculate the Ly$\alpha$ 2D autocorrelation from the observations, as the quasar sample does not have a high enough sky density to probe the on-sky projected profile.

## 7 DISCUSSION

We have combined the power of the VLRS at large spatial scales with the statistical power of the Keck sample at smaller scales. Crighton et al. (2011) included the Keck data in the LBG–Ly$\alpha$ cross-correlation function by simply using an error-weighted combination of the Keck and VLRS correlation functions. Our aim here was to combine the two surveys for 2D, $\xi(\sigma, \pi)$ correlation function analyses at the deeper level of the Ly$\alpha$ fluxes and LBG positions. We therefore included 940 $2.67 \le z \le 3.25$ LBGs from the Steidel et al. (2003) Keck samples. We also re-reduced six high-resolution spectra of the quasars in these fields from the ESO and Keck archives. With $\approx 3000$ galaxies, the combined VLRS and Keck surveys covering the widest range of spatial scales are ideal to study the dynamical relationship between galaxies and the IGM at $z \approx 3$.

We have also incorporated the GIMIC SPH simulation into our analysis, in order to aid the interpretation of the correlation function results. GIMIC was used to create synthetic Ly$\alpha$ spectra and galaxies. We study both galaxy clustering and the relationship between gas and galaxies via the auto- and cross-correlation functions in both 1D and 2D.

We have compared the simulated galaxy–galaxy results in real and redshift space. The simulated galaxy autocorrelation functions, $\xi(r)$ and $\xi(s)$ (i.e. in real and redshift space), are consistent with power laws at scales of $r \gtrsim 2\,h^{-1}$ Mpc. At small distances ($r \lesssim 1\,h^{-1}$ Mpc), the LBG–LBG $\xi(s)$ tends to have lower clustering than $\xi(r)$ in real space, while at larger scales the LBG–LBG $\xi(s)$ results have higher clustering. Qualitatively, this is as expected from 'finger-of-God' effects at sub $1\,h^{-1}$ Mpc scales and dynamical infall at larger scales, characterized by the 'Kaiser boost'. Quantitatively, the large-scale Kaiser boost for the galaxies is marginally lower than predicted based on the galaxy bias, but only by $\approx 1$–$2\sigma$. At smaller scales, the peculiar velocity dispersion measured in the simulation overestimates the difference between real- and redshift-space correlation functions. Similar results have been found by Taruya, Nishimichi & Saito (2010), who found that at high redshift fitting finger-of-God damping terms, as we do here, tended to underestimate the peculiar velocity dispersion predicted by linear theory. Certainly, a 'local' velocity dispersion measured relative to galaxy pairs with separations $<1\,h^{-1}$ Mpc produces improved agreement.

From the simulated galaxy 2D autocorrelation function, $\xi(\sigma, \pi)$, we find values for the galaxy infall parameter, $\beta_{\rm gal}$, consistent within $\approx 1$–$1.5\sigma$ with what would be expected from the measured galaxy bias. The same is seen for the pairwise velocity dispersion. Overall, our RSD model is successful in retrieving the properties of the galaxy velocity field when applied to the clustering measurements from the simulation, whilst conversely the simulation is shown to reproduce a realistic galaxy velocity field well.

Following the galaxy autocorrelation analysis, we performed an analysis of the galaxy–gas cross-correlation. We first analysed the LBG–Ly$\alpha$ $\langle T(s) \rangle$ 1D cross-correlation function as calculated directly from quasar sightline spectra and LBG positions from the VLRS and Keck surveys. We have re-analysed a subset of six fields from the eight used in the work of A03 and found good agreement, observing the small-scale upturn reported in the original work. Our VLRS results on the other hand, agree with those of A05 (and Rakic et al. 2012), rather than those of A03, i.e. a continuous decrease in flux transmissivity around the LBG with no evidence for a spike in transmissivity. Crighton et al. (2011) noted that such a spike could still be present but smoothed away by the errors in the LBG velocities, but this now seems unlikely given the results presented here and those of A05 and Rakic et al. (2012).

The inclusion of the gas component along with the galaxies allows us to investigate the effects of gaseous infall on the galaxy–gas distribution. Fitting an RSD model to the observed VLRS+Keck LBG–Ly$\alpha$ $\xi(\sigma, \pi)$, we found best-fitting parameters of $\beta_{\rm Ly\alpha} = 0.33^{+0.33}_{-0.23}$ and $\langle w_z^2 \rangle^{1/2} = 190 \pm 90\,{\rm km\,s^{-1}}$. The large-scale infall measurement is significantly lower than that predicted by McDonald (2003), i.e. $\beta_{\rm Ly\alpha} = 1.3 \pm 0.3$, whilst the velocity dispersion measurement is consistent (although lower by $\sim 1\sigma$) with the velocity errors in our galaxy redshifts. Interestingly, the second point here leaves little room for any intrinsic velocity dispersion between the gas and galaxies at small scales. We see similar results when analysing the simulated galaxy–Ly$\alpha$ cross-correlation. Again, we find $\beta_{\rm Ly\alpha} \approx 0.3$, whilst the velocity dispersion is measured to be somewhat lower than we might expect for the gas–galaxy velocity dispersion based on their directly measured individual velocity profiles (i.e. from Fig. 6). Indeed, these small measurements of the galaxy–gas velocity dispersion in both our observations and simulations may be indicative of highly coherent motion between gas and galaxies at small scales.

From the Ly$\alpha$ autocorrelations $\xi(r)$ and $\xi(\sigma, \pi)$, we see similar results, again with small differences between real and redshift

space. At small scales, the velocity dispersion needed to fit the simulated $\xi(s)$ is less than measured directly in the simulation, although this may be partly explained by the intrinsic width of the Ly$\alpha$ lines contributing artificial autocorrelation below separations of $\lesssim 0.7\,h^{-1}$ Mpc. At larger scales, the value of $\xi(s)/\xi(r)$ gives $\beta_{\rm Ly\alpha} = 0.4 \pm 0.16$ rather than the range given by McDonald (2003), $\beta_{\rm Ly\alpha} \approx 1$–$1.6$ (but entirely consistent with the results from the GIMIC cross-correlation).

At larger scales, one possibility to explain the low gas infall rate may be the presence of feedback in the GIMIC simulations. Galaxy-wide winds powered with initial velocities of $600\,{\rm km\,s^{-1}}$ are invoked in the GIMIC simulations and this is a significant amount, since this corresponds to $6\,h^{-1}$ Mpc. These winds are modelled by each star particle that forms, imparting a randomly directed $600\,{\rm km\,s^{-1}}$ kick to four of its gas particle neighbours. It is possible that this outflow of the gas could cancel out some of the expected gravitational infall, particularly in the neighbourhood of a galaxy. However, it remains to be seen whether enough gas particles are outflowing to explain the lack of infall in the gas cross- or autocorrelation functions. If the effects of gas outflow were detectable in the gas dynamics, this could be a powerful probe, since there is no evidence of feedback from any spike in transmission due to lower neutral gas density close to the galaxy.

Studies by Rakic et al. (2012) and Rakic et al. (2013) presented the LBG–H I cross-correlation at $z \sim 2.4$ with observations and simulations respectively. In both cases, the authors report a significant measurement of RSD, showing evidence for both small-scale peculiar velocity effects and large-scale bulk motion of gas infalling on to observed and simulated galaxies. Rakic et al. (2013) find that, in terms of the reported large-scale 'flattening', the observations of Rakic et al. (2012) are consistent with the simulation results for galaxy samples selected with minimum halo masses of $\log(M_{\rm min}/M_{\odot}) = 11.6 \pm 0.2$. This is consistent with the halo masses (measured from galaxy clustering) of Trainor & Steidel (2012), but is significantly higher than the halo masses of the galaxy samples used here (and those of Bielby et al. 2013, A03 and A05). We are unable to probe this larger halo-mass constraint given the size limitations of GIMIC; however, we note that our GIMIC 2D cross-correlation results appear qualitatively consistent with the results of Rakic et al. (2013) at lower minimum halo masses. Additionally, Rakic et al. (2013) compared their measurements for different feedback prescriptions, finding that including AGNs weakened the absorption by H I (within $\sim 1$ Mpc), whilst increasing the wind mass-loading increased the measured absorption. The authors do not make any quantitive analysis of the effect of increasing the wind mass-loading on the presence of large-scale infall in the cross-correlation analysis. However inspecting their fig. 4, it is evident that there is indeed some movement in the large-scale measurement of the gas distribution when the wind mass-loading is increased (i.e. comparing the 'REF' model result with the 'WML4' result). This provides some additional motivation for the supposition that SNe-driven winds could affect our measurement of $\beta_{\rm Ly\alpha}$. An important test of this will be to apply our RSD modelling to a range of simulation runs incorporating different feedback prescriptions.

Rakic et al. (2012) also investigate the effect of small-scale random peculiar velocities on their observed cross-correlation, finding evidence for peculiar velocities between gas and galaxies of $\sim 240\,{\rm km\,s^{-1}}$. Such a large peculiar velocity is not apparent in the simulation results of Rakic et al. (2013) and neither is it in our simulation results. Our observations give a measurement of the velocity dispersion consistent with that reported by Rakic et al.

(2012); however, this is largely dominated by galaxy redshift errors, as we have discussed.

The Rakic et al. (2012) results have since been further developed by Turner et al. (2014), in whose work increased numbers of the galaxy sample have been observed using the MOSFIRE instrument at the Keck Observatory (improving redshift accuracies). Turner et al. (2014) report consistent results with Rakic et al. (2012) for the galaxy–H I cross-correlation, seeing the same finger-of-God and large-scale infall effects with their improved redshift errors. It is difficult to make a quantitive comparison between our results and those of Rakic et al. (2012, 2013) and Turner et al. (2014), however, given the different measurements made. Qualitatively, these complementary studies are consistent with the work presented here in identifying the presence of both large-scale infall and small-scale peculiar velocity effects in the H I gas around $z = 2$–3 star-forming galaxies.

A more direct comparison can be made with the results of Slosar et al. (2011), who measure the $\beta_{Ly\alpha}$ parameter from the autocorrelation of the Ly$\alpha$ forest in BOSS quasar spectra. They find a range $0.44 < \beta_{Ly\alpha} < 1.20$, at central redshift $z = 2.25$. This large range is, however, consistent at the $1\sigma$ level with all other results considered, i.e. the VLRS+Keck observations, the GIMIC simulation results and the theoretical prediction from McDonald (2003). Interestingly, though, assuming $\beta_{Ly\alpha}$ behaves as $\beta_{gal}$ it should then decrease with increasing redshift and we would expect the $z = 3$ result to be marginally lower than $\beta_{Ly\alpha}$ at $z = 2.25$, which is what we find in our study.

## 8 CONCLUSIONS

We have analysed the interaction between galaxies and the IGM using a large sample of $z \sim 3$ LBGs, in combination with spectroscopic observations of background quasars. In addition to the observational data, we employ the SPH GIMIC simulation to analyse the clustering of gas and galaxies.

(1) We analyse the autocorrelation of simulated galaxies in the GIMIC simulation using two samples: $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ and $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$. The $M_* > 10^9 \, h^{-1} \, \mathrm{M}_\odot$ sample was chosen to match the clustering amplitude of observed LBGs, while the $M_* > 10^8 \, h^{-1} \, \mathrm{M}_\odot$ sample provides a comparison set with higher numbers and hence better statistics. In the simulated data, the difference between the real- and redshift-space correlation functions is too small to be explained self-consistently by the measured peculiar velocity distribution. We suggest that this is the consequence of a scale dependence in the measurement of the peculiar motions and that the peculiar motions taken within $\lesssim 1 \, h^{-1}$ Mpc give a more consistent result.

(2) We have checked for the existence of a transmission spike near star-forming galaxies in the data and GIMIC simulations, which could be indicative of the effects of star-formation feedback on the IGM. For the data, we combined the full VLRS and Keck LBG–Ly$\alpha$ data sets to study both $\xi(r)$ and $\xi(\sigma, \pi)$ and the LBG–Ly$\alpha$ correlation functions. We find no evidence for a transmission spike at small scales and instead find that the gas transmissivity drops monotonically towards the galaxy, consistent with the density of neutral gas rising towards the galaxy position. Although the simulation transmission rises when LBG velocity errors are taken into account, the simulated and observational results remain in good statistical agreement.

(3) The redshift-space galaxy–Ly$\alpha$ cross-correlation function in the simulation is close to the real-space correlation function and to

some extent this is predicted from linear theory applied to the Ly$\alpha$ forest flux, which has a non-linear relation to optical depth and thus implies lower rates of dynamical infall of gas into galaxies than would otherwise apply. We have also considered whether galaxy-wide outflows may be cancelling out the infall effect.

(4) The observed Ly$\alpha$ autocorrelation function is also consistent with the simulation. At small scales the difference between real- and redshift-space correlation functions in the simulation is again less than predicted, given the peculiar velocity distribution. At larger scales, we measure the effects of dynamical infall and find them to be less than predicted, based on the simulations of McDonald (2003). This may be a residual effect from gas outflows cancelling out the effects of dynamical infall.

(5) In the simulations, both gas and galaxies show evidence of a strong bulk motion. This bulk motion is undetectable by observable correlation functions but may have a connection with the local coherence needed to explain why the distribution of peculiar velocities overestimates the finger-of-God effect.

## 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., 2005a, ApJ, 629, 636
Adelberger K. L., Steidel C. C., Pettini M., Shapley A. E., Reddy N. A., Erb D. K., 2005b, ApJ, 619, 697
Aguirre A., Schaye J., Hernquist L., Kay S., Springel V., Theuns T., 2005, ApJ, 620, L13
Altay G., Theuns T., Schaye J., Crighton N. H. M., Dalla Vecchia C., 2011, ApJ, 737, L37
Bielby R. M. et al., 2011, MNRAS, 414, 2
Bielby R. et al., 2013, MNRAS, 430, 425
Bruscoli M., Ferrara A., Marri S., Schneider R., Maselli A., Rollinde E., Aracil B., 2003, MNRAS, 343, L41
Cen R., Ostriker J. P., 1999, ApJ, 519, L109
Crain R. A. et al., 2009, MNRAS, 399, 1773
Crighton N. H. M. et al., 2011, MNRAS, 414, 28
Croft R. A. C., Hernquist L., Springel V., Westover M., White M., 2002, ApJ, 580, 634
da Ângela J., Outram P. J., Shanks T., Boyle B. J., Croom S. M., Loaring N. S., Miller L., Smith R. J., 2005, MNRAS, 360, 1040
Dalla Vecchia C., Schaye J., 2008, MNRAS, 387, 1431
Davis M., Peebles P. J. E., 1983, ApJ, 267, 465
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dekel A. et al., 2009, Nature, 457, 451
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekker H., D'Odorico S., Kaufer A., Delabre B., Kotzlowski H., 2000, in Iye M., Moorwood A. F., eds, Proc. SPIE Conf. Ser. Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. SPIE, Bellingham, p. 534
Desjacques V., Nusser A., Haehnelt M. G., Stoehr F., 2004, MNRAS, 350, 879
Desjacques V., Haehnelt M. G., Nusser A., 2006, MNRAS, 367, L74

Dolag K., Borgani S., Murante G., Springel V., 2009, MNRAS, 399, 497

Faucher-Giguère C.-A., Prochaska J. X., Lidz A., Hernquist L., Zaldarriaga M., 2008, ApJ, 681, 831

Font A. S., McCarthy I. G., Crain R. A., Theuns T., Schaye J., Wiersma R. P. C., Dalla Vecchia C., 2011, MNRAS, 416, 2802

Frenk C. S., Evrard A. E., White S. D. M., Summers F. J., 1996, ApJ, 472, 460

Hamilton A. J. S., 1992, ApJ, 385, L5

Hawkins E. et al., 2003, MNRAS, 346, 78

Hildebrandt H., Pielorz J., Erben T., van Waerbeke L., Simon P., Capak P., 2009, A&A, 498, 725

Hirschmann M. et al., 2013, MNRAS, 436, 2929

Kaiser N., 1987, MNRAS, 227, 1

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

Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, MNRAS, 395, 160

Kollmeier J. A., Weinberg D. H., Davé R., Katz N., 2003, ApJ, 594, 75

Le Fèvre O. et al., 2003, in Iye M., Moorwood A. F. M., eds, Proc. SPIE Conf. Ser. Vol. 4841, Instrument Design and Performance for Optical/Infrared Ground-based Telescopes. SPIE, Bellingham, p. 1670

Matsubara T., Suto Y., 1996, ApJ, 470, L1

McCarthy I. G., Font A. S., Crain R. A., Deason A. J., Schaye J., Theuns T., 2012a, MNRAS, 420, 2245

McCarthy I. G., Schaye J., Font A. S., Theuns T., Frenk C. S., Crain R. A., Dalla Vecchia C., 2012b, MNRAS, 427, 379

McDonald P., 2003, ApJ, 585, 34

McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1

Mountrichas G., Sawangwit U., Shanks T., 2009, MNRAS, 398, 971

Navarro J. F. et al., 2004, MNRAS, 349, 1039

Oke J. B. et al., 1995, PASP, 107, 375

Oppenheimer B. D., Davé R., 2006, MNRAS, 373, 1265

Oppenheimer B. D., Davé R., 2008, MNRAS, 387, 577

Pettini M., Shapley A. E., Steidel C. C., Cuby J.-G., Dickinson M., Moorwood A. F. M., Adelberger K. L., Giavalisco M., 2001, ApJ, 554, 981

Planck Collaboration et al., 2013, preprint (arXiv:1303.5076)

Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14

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

Rakic O., Schaye J., Steidel C. C., Booth C. M., Dalla Vecchia C., Rudie G. C., 2013, MNRAS, 433, 3103

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

Scannapieco C. et al., 2012, MNRAS, 423, 1726

Schaye J. et al., 2010, MNRAS, 402, 1536

Schaye J., Dalla Vecchia C., 2008, MNRAS, 383, 1210

Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, ApJ, 588, 65

Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, ApJ, 626, 698

Slosar A. et al., 2011, J. Cosmology Astropart. Phys., 9, 1

Springel V. et al., 2005, Nature, 435, 629

Springel V., 2005, MNRAS, 364, 1105

Springel V., Hernquist L., 2003, MNRAS, 339, 289

Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726

Steidel C. C., Adelberger K. L., Shapley A. E., Pettini M., Dickinson M., Giavalisco M., 2003, ApJ, 592, 728

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

Taruya A., Nishimichi T., Saito S., 2010, Phys. Rev. D, 82, 063522

Theuns T., Leonard A., Efstathiou G., Pearce F. R., Thomas P. A., 1998, MNRAS, 301, 478

Theuns T., Viel M., Kay S., Schaye J., Carswell R. F., Tzanavaris P., 2002, ApJ, 578, L5

Trainor R. F., Steidel C. C., 2012, ApJ, 752, 39

Turner M. L., Schaye J., Steidel C. C., Rudie G. C., Strom A. L., 2014, preprint (arXiv:1403.0942)

van de Voort F., Schaye J., Altay G., Theuns T., 2012, MNRAS, 421, 2809

Vernet J. et al., 2011, A&A, 536, A105

Vogt S. S. et al., 1994, in Crawford D. L., Craine E. R., eds, Proc. SPIE Conf. Ser. Vol. 2198, Instrumentation in Astronomy VIII. SPIE, Bellingham, p. 362

Weil M. L., Eke V. R., Efstathiou G., 1998, MNRAS, 300, 773

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

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, MNRAS, 399, 574

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