The correlation of high-redshift galaxies with the thermal Sunyaev–Zel'dovich effect traces reionization

Eric J. Baxter,[1,2,3,4]★ Lewis Weinberger [2,3] Martin Haehnelt,[2,3] Vid Iršič,[5,3] Girish Kulkarni [6] Shivam Pandey[7] and Anirban Roy[8]

[1]*Institute for Astronomy, University of Hawai'i, 2680 Woodlawn Drive, Honolulu, HI 96822, USA*
[2]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[3]*Kavli Institute for Cosmology, University of Cambridge, Cambridge CB3 0HA, UK*
[4]*Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, UK*
[5]*Cavendish Laboratory, University of Cambridge, 19 J. J. Thomson Ave., Cambridge CB3 0HE, UK*
[6]*Department of Theoretical Physics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[7]*Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA 19104, USA*
[8]*Department of Astronomy, Cornell University, Ithaca, NY 14853, USA*

ABSTRACT

We explore a potential new probe of reionization: the cross-correlation of high-redshift galaxies with maps of the thermal Sunyaev–Zel'dovich (tSZ) effect. We consider two types of high-redshift galaxies: Lyman break galaxies (LBGs) and Lyman-α emitters (LAEs). LBGs and LAEs will be detected in large numbers at high redshift ($z \approx 4$–7) by ongoing and future surveys. We consider a future LBG sample from The Rubin Observatory Legacy Survey of Space and Time, and a selection of LAEs modelled after the Subaru SILVERRUSH programme, but covering a much larger sky fraction. The tSZ effect is sensitive to a line-of-sight integral of the ionized gas pressure, and is measured across large patches of sky using multifrequency cosmic microwave background (CMB) surveys. We consider forecast tSZ maps from CMB Stage 4 and more futuristic observations. Using a suite of hydrodynamical simulations, we show that LBGs and LAEs are correlated with the tSZ signal from reionization. The cross-spectra between LBGs/LAEs with tSZ maps contain information about the reionization history of the Universe, such as the distribution of bubble sizes, and could be used to directly measure the timing of reionization. The amplitude of the signal is small, however, and its detectability is hindered by low-redshift contributions to tSZ maps and by instrumental noise. If the low-redshift contribution to the observed tSZ signal is suppressed by masking of massive haloes, a combination of overlapping futuristic CMB and galaxy surveys could probe this signal.

Key words: dark ages, reionization, first stars – large-scale structure of Universe – cosmology: theory.

1 INTRODUCTION

Reionization impacts the observed cosmic microwave background (CMB) in several ways. For one, CMB photons may Thompson scatter with the free electrons released by reionization. This process impacts the observed temperature and polarization power spectra, enabling measurement of the optical depth to the last scattering surface, τ. A constraint on τ in turn constrains the timing of reionization, since an earlier onset will increase the volume of ionized gas that CMB photons must traverse, thereby increasing τ (for a review, see Reichardt 2016). CMB observations also constrain reionization via the kinematic Sunyaev–Zel'dovich (kSZ) effect (Sunyaev & Zeldovich 1980). The kSZ is a Doppler shift imparted to CMB photons by inverse Compton scattering with ionized gas that has bulk velocity relative to the CMB rest frame (for a review, see Birkinshaw 1999). The inhomogeneously ionized Universe during reionization leads to a 'patchy' kSZ signal that can be observed by

CMB experiments (Gruzinov & Hu 1998). Measurements of this patchy signal constrain the duration of reionization (e.g. Reichardt et al. 2020).

Reionization should also leave an impact in the CMB via the *thermal* Sunyaev–Zel'dovich (tSZ) effect (Sunyaev & Zeldovich 1972), which results from inverse Compton scattering of CMB photons with hot ionized gas. This process leads to a spectral distortion in the CMB that is sensitive to Compton-y, which is in turn proportional to a line-of-sight integral of the ionized gas pressure:

$$y = \frac{\sigma_T}{m_e c^2} \int_0^\infty \mathrm{d}l \, P_e(l), \tag{1}$$

where $P_e(l)$ is the electron pressure and l is the line-of-sight distance. The same bubbles of warm ionized gas that lead to the patchy kSZ signal should also lead to an inhomogenous y signal. Since y scales like $k_B T/m_e c^2$, where T is the gas temperature, while the kSZ signal scales like v/c, where v is the gas bulk velocity, the relative amplitude of the tSZ signal to the kSZ signal is roughly $T k_B/m_e c v$. The temperature inside the reionized bubbles of gas is roughly $T \sim 10^4$ K, and a characteristic velocity is $v \sim 200 \, \mathrm{km \, s^{-1}}$.

Substituting these values, the tSZ signal from reionization is expected to be about two orders of magnitude smaller than the kinematic signal. However, the *y* signal carries additional information about the thermodynamics of the gas that could in principle be used to constrain reionization models. Furthermore, unlike the kSZ, the tSZ leads to a distinct spectral distortion in the CMB, making it more straightforward to estimate from multifrequency CMB observations than the blackbody-preserving kSZ signal.

The *y* signal is sensitive to all gas between the observer and the last scattering surface. Because the hot gas in haloes at low redshift makes a dominant contribution to the integrated pressure, measurements of the *y* power spectrum are essentially only sensitive to the low-redshift gas that lives in the most massive dark matter haloes (e.g. Makiya, Ando & Komatsu 2018). Extracting the high-redshift contribution to the *y* power spectrum would require modelling (and subtracting) the low-redshift contribution with extremely high accuracy.

However, by cross-correlating maps of Compton-*y* with a tracer of the large-scale structure at high redshift, such as galaxies, one can effectively isolate the contributions to *y* from the high-redshift Universe. In this case, the low-redshift contribution acts solely as a noise source. A detection of this correlation as a function of redshift has the potential to directly probe how the Universe reionized over cosmic time.

One way to select large populations of high-redshift galaxies for this purpose is via wide-field photometric surveys. Ongoing (e.g. the GOLDRUSH catalogue from HSC; Ono et al. 2018) and future [e.g. the Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST),[1] Ivezić et al. 2019; WFIRST, Spergel et al. 2015; EUCLID, Laureijs et al. 2011] deep imaging surveys can employ so-called dropout methods to select high-redshift galaxies. As the Lyman break feature in a galaxy spectrum passes through the optical filter bands, the galaxy will disappear (i.e. dropout) in the bluer bands. Simple colour and flux cuts can be used to select populations of these Lyman break galaxies (LBGs) localized in fairly narrow redshift intervals. For instance, *g*-band dropouts are localized at around $z \sim 4$, *r*-band dropouts are localized around $z \sim 5$, etc. Large samples of dropout-selected galaxies are expected from LSST and other future imaging surveys (Wilson & White 2019).

Another class of galaxies that can be relatively easily identified at high redshift are Lyman-α emitters (LAEs). Narrow-band searches can be used to identify the redshifted Lyman-α emission from these galaxies out to $z \sim 7$. While the largest samples of LAEs to date have been derived from fairly small area surveys – such as the $\sim 20 \deg^2$ SILVERRUSH survey using the Subaru telescope (Ouchi et al. 2018) – future samples may be obtained from much larger sky areas using e.g. SPHEREx (Doré et al. 2014).

In this work, we explore the cross-correlation signal between high-redshift galaxies (both LBGs and LAEs) and Compton-*y* using the Sherwood simulation suite (Bolton et al. 2017), a set of IGM-focused hydrodynamical simulations. We show that the galaxy–*y* correlation provides a distinctive signal of reionization. We then determine whether future surveys have a hope of detecting this signal.

Several recent works (e.g. Sobacchi, Mesinger & Greig 2016; Kubota et al. 2018) have considered the related possibility of probing reionization with measurement of the correlation between high-redshift galaxies, such as LAEs, and maps of the redshifted 21-cm signal from neutral hydrogen. The correlation considered here – between similar galaxies and *y* – is complementary to the 21-cm correlation, since it provides additional information about the

gas pressure in the ionized bubbles. Additionally, the estimation of the *y* and 21-cm signals from data must contend with different observational challenges. Brightness temperature fluctuations due to galactic synchrotron and extragalactic point sources are orders of magnitude larger than the expected 21-cm signal (e.g. Shaver et al. 1999). In the case of *y*, the main observational challenges are instrument noise and contributions to *y* from low-redshift structure, as we discuss in more detail below.

The paper is organized as follows. In Section 2, we describe the simulations used to compute the cross-correlation between galaxies and Compton-*y*; in Section 3, we described the methodology for forecasting constraints on these cross-correlations from future surveys; our results are presented in Section 4, and we conclude in Section 5. In Appendix A, we explore how our results are impacted by the feedback modelling within our simulations, while in Appendix B we provide simple analytical estimates for the amplitude of the *y* signal from reionized bubbles.

2 SIMULATIONS

We employ the same simulation set-up as in Roy et al. (2021), using the Sherwood simulation suite (Bolton et al. 2017) to model the LBG, LAE, and *y* signals. This suite of cosmological hydrodynamical simulations was designed to capture the evolution of the low-density inter-galactic medium (IGM). We employ the delayed reionization history that has recently been found to explain Ly α forest opacity fluctuations (Kulkarni et al. 2019; Nasir & D'Aloisio 2020; Keating et al. 2020), the high-redshift LAE luminosity function evolution (Weinberger, Haehnelt & Kulkarni 2019), and the low Thomson scattering optical depth to the CMB (Planck Collaboration VI 2020).

2.1 Reionizing the Sherwood simulations

The Sherwood suite was run using the P-GADGET-3 smoothed-particle hydrodynamics (SPH) code (Springel, Yoshida & White 2001; Springel 2005), with a range of simulation box sizes and resolutions. In this work, we make use of the largest volume which has a side length of $L = 160$ cMpc h^{-1}[2] and particle number $N = 2 \times 2048^3$, giving a dark matter particle mass of $M_{\rm DM} = 3.44 \times 10^7 \, {\rm M}_\odot \, h^{-1}$. The cosmological parameters were derived from the Planck Collaboration XVI (2014) results, including $h = 0.678$, $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_{\rm b} = 0.0482$, $\sigma_8 = 0.829$, $n = 0.961$, and $Y_{\rm He} = 0.24$. To improve computational efficiency, the QUICK_LYA star formation prescription (Viel, Haehnelt & Springel 2004) was employed, in which cold dense gas ($T < 10^5$ K, $\Delta > 10^3$) is converted into star particles (see Appendix A for discussion of the impact of this prescription on our results). From this simulation, we use the SPH kernel to interpolate the hydrodynamic quantities on to a uniform grid at our target redshifts of interest $z = 5.000, 5.756, 5.946, 6.604, 6.860, 7.444$. These redshifts overlap with the broad-band *i* and *r* filters (which can be used to identify high-redshift LBGs via dropout selection), and the narrow-band NB816, NB921, NB973, and NB101 filters (which can be used to identify high-redshift LAEs).

We employ the late reionization history of Kulkarni et al. (2019) by post-processing the simulations with the radiative transfer code ATON (for further details of the code, see Aubert & Teyssier 2010, 2008; for details of the simulation set-up, see Kulkarni et al. 2019). We note that the shape and evolution of the cross-correlation signal

---

[2] We use the prefix c to denote comoving units, while p indicates proper units.
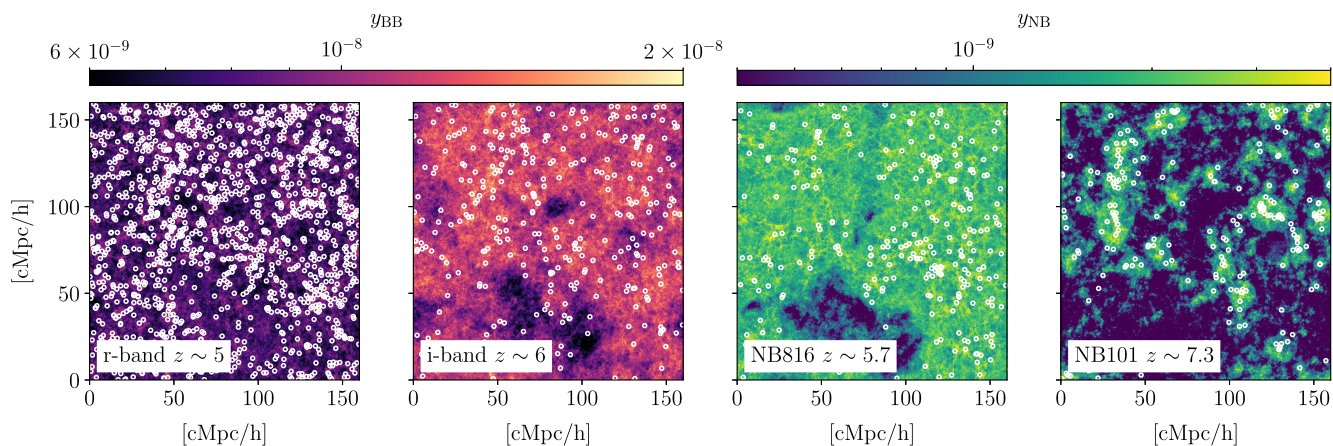
Figure 1. Simulated galaxy populations and the contribution to the Compton-y signal from within the corresponding galaxy survey band. From left to right, we show the slices at $z = 5.000$, 5.946, 5.756, and 7.444, with mock LBG and LAE positions indicated by white circles. The left two panels show the simulated y-signal from integrating across the mock broad-bands, y_{BB}, corresponding to r- and i-band dropout galaxy selections; the mock LBG population are overlayed as white circles. The right-hand panels similarly show the simulated y-signal from integrating across the mock NB816 and NB101 narrow bands, y_{NB}, with the LAE population overlayed as white circles. Both galaxy populations are highly spatially correlated with the y-signal.
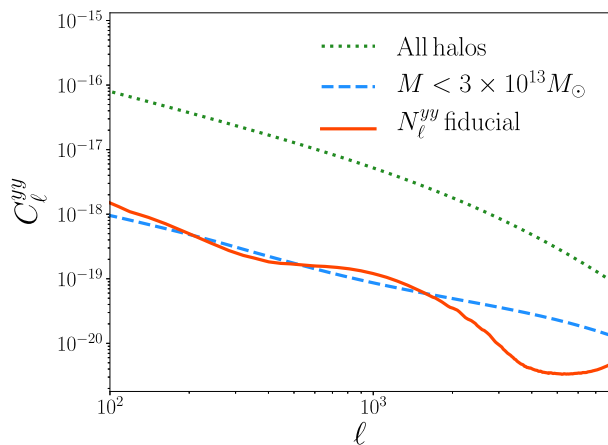

Figure 2. The Compton-y autospectrum before (green dotted) and after (blue dashed) masking haloes with $M > 3 \times 10^{13}$ M$_\odot$. Masking massive haloes significantly reduces the y autospectrum, which acts as a source of noise for the measurement of the cross-correlation between high-redshift galaxies and y. Also shown is the fiducial noise level assumed for forecasts (red solid curve), which is a factor of 10 lower than that expected for CMB-S4.

is dependent on the reionization history (which controls the ionized bubble distribution; see Appendix B) and hence by fixing this choice we are forecasting only for this specific case. As discussed further in Roy et al. (2021) and Weinberger, Kulkarni & Haehnelt (2020), our chosen reionization history is consistent with existing observations of the Lyman-α forest (Kulkarni et al. 2019; Nasir & D'Aloisio 2020; Keating et al. 2020), LAE populations (Weinberger et al. 2019), and the CMB optical depth (Planck Collaboration VI 2020). We refer the interested reader to fig. 1 of Kulkarni et al. (2019) for further details of this reionization history.

To push our forecasts to scales larger than our fiducial 160 cMpc h^{-1} simulation volume, we also employ a seminumerical simulation with a side length of 1 cGpc (Roy et al. 2021), run using 21cmFast (Mesinger, Furlanetto & Cen 2011). This simulation lacks the spatial resolution to accurately model the smaller scale correlations between y and the galaxy populations, however it is effective for predicting the large-scale modes. We post-process this

simulation with the same reionization history as the Sherwood simulation, using the calibrated excursion set method of Choudhury et al. (2015). For further implementation details and a comparison of these simulations, we refer the reader to Roy et al. (2021).

While the bulk of our analysis focuses on results from numerical simulations, in Appendix B we provide analytical estimates for the amplitude of the signal that are in good agreement with the simulation results.

2.2 Modelling the *y* signal

It is useful to consider the contribution to the total y from structure between redshifts z_{min} and z_{max}. Reframing equation (1) as an integral over redshift (here using the *comoving* electron pressure), we have

$$y(z_{min}, z_{max}) = \frac{\sigma_T}{m_e c^2} \int_{z_{min}}^{z_{max}} dz\, (1+z)^2\, \frac{d\chi}{dz}\, P_e(z)$$
$$\equiv \int_{z_{min}}^{z_{max}} dz\, W_y(z)\, P_e(z), \tag{2}$$

where χ is comoving distance, and the window function is given by,

$$W_y(z) \equiv \frac{\sigma_T}{m_e c^2} (1+z)^2\, \frac{d\chi}{dz}. \tag{3}$$

The y-signal measured by CMB experiments is sensitive to the full integral of the ionized gas pressure from to $z_{min} = 0$ to $z_{max} = z_{CMB} \sim 1100$ (see equation 1). However, the cross-correlation with a galaxy survey will pick out the contribution to y from within the survey volume. In practice, we perform the integral in equation (2) across the redshift limits of the galaxy bands, and use the notation y_{BB} or y_{NB} (broad-band and narrow-band, respectively) for this 'band-contribution' to the observed (total) y. This procedure will capture the contributions to y relevant for estimation of the cross-correlation signals. Of course, the redshift distributions of the selected galaxies will not be exactly described by top-hat functions, so equation (2) is an approximation. Furthermore, in actual data, there is the possibility of low-redshift interlopers contaminating the galaxy samples (e.g. Dunlop 2013). Given our focus on idealized forecasts, we postpone a careful consideration of these issues to future work. Finally, equation (2) ignores contributions to y that may be correlated
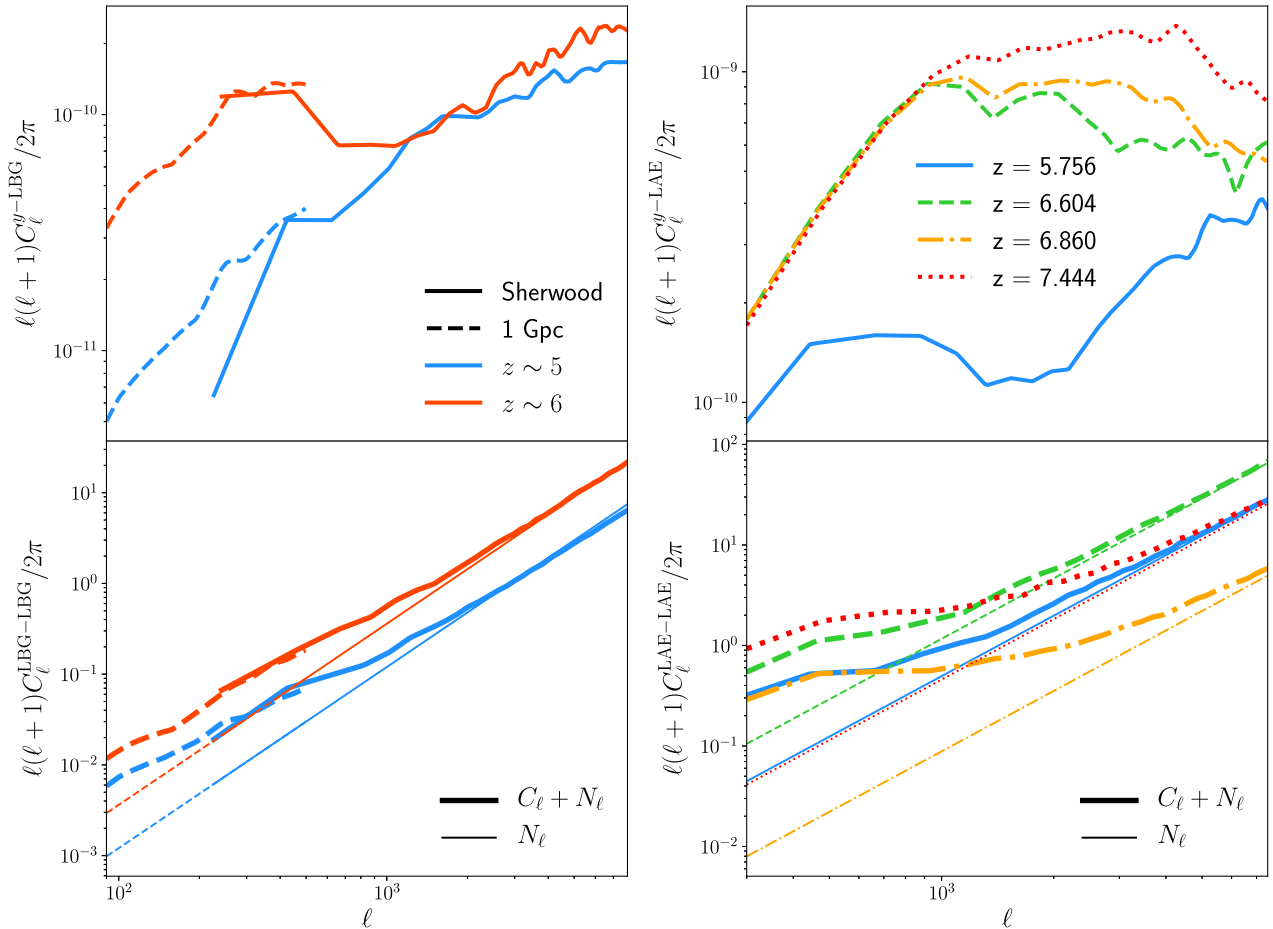
Figure 3. The galaxy–y cross-spectra (top panels) and galaxy clustering autospectra (bottom panels) computed from the simulations. The left-hand panels show the results for LBGs, while the right-hand panels show the results for LAEs. For the LBG results, we show power spectra computed for the 160 cMpc h^{-1} Sherwood simulation with solid lines, and power spectra for the 1 cGpc simulation with dashed lines. In the bottom panels, the thick curves correspond to the total (signal and noise) galaxy autopower, while the thin lines correspond to the shot noise for each galaxy sample.

with the selected galaxies, but that extend outside of the bands considered. Given that in all cases, the bands are several times wider than the typical correlation length of the galaxy and y fields, we expect these contributions to be small. Our choice to ignore these contributions is motivated by the fact that uncorrelated y in the simulated y maps will add noise to the correlation measurements.

After interpolating the SPH particles and running the post-processing radiative transfer, we have grids of x_e, n_H, and T that can be used to derive P_e. The simulated contribution to the y-signal from broad-band and narrow-band survey volumes is shown in Fig. 1. These maps were created by integrating equation (2) across the redshift range of the band. We note that we do not account for the light-cone effect here, which may introduce some inaccuracy for the broad-band filters that span $\Delta z \gtrsim 1$.

2.3 Modelling the LBG and LAE populations

We use the galaxy modelling described in Weinberger et al. (2019). We will briefly summarise the details of this modelling here, and refer the interested reader to that work for further information.

We start with the dark matter halo population of the simulation, which is found on-the-fly with a Friends-of-Friends finder. At $z = 5$ the minimum and maximum halo masses are $M_h^{\min} = 2.3 \times 10^8$

$M_\odot\, h^{-1}$ and $M_h^{\max} = 4.6 \times 10^{12}\ M_\odot\, h^{-1}$, respectively, while at $z = 7.444$ the maximum mass is $M_h^{\max} = 2.7 \times 10^{12}\ M_\odot\, h^{-1}$. This halo population is abundance matched to the ultraviolet (UV) luminosity function from Bouwens et al. (2015) at $z \sim 6$ with the duty cycle formalism from Trenti et al. (2010) in order to create a mock LBG population. The resulting mapping from halo mass to broad-band UV luminosity, $L_{UV}(M_h)$, can be applied to the halo population at other redshifts to recover the observed evolution in the luminosity function.

To generate the LAE population, we find the subset of this LBG population with observable Ly α luminosities and equivalent widths. The Ly α properties of the LBG population are modelled statistically using the equivalent width distribution model of Dijkstra & Wyithe (2012). We calculate the IGM transmission along single sightlines through the simulated IGM to each LAE in order to account for the effect of reionization[3] (for further details, see Weinberger et al. 2018).

---

[3]We note that we only model the LAE population in the 160 cMpc h^{-1} Sherwood simulation and not the 1 cGpc simulation, as the Ly α transmission calculation is sensitive to how well the simulation can resolve the small-scale self-shielded neutral hydrogen around haloes.
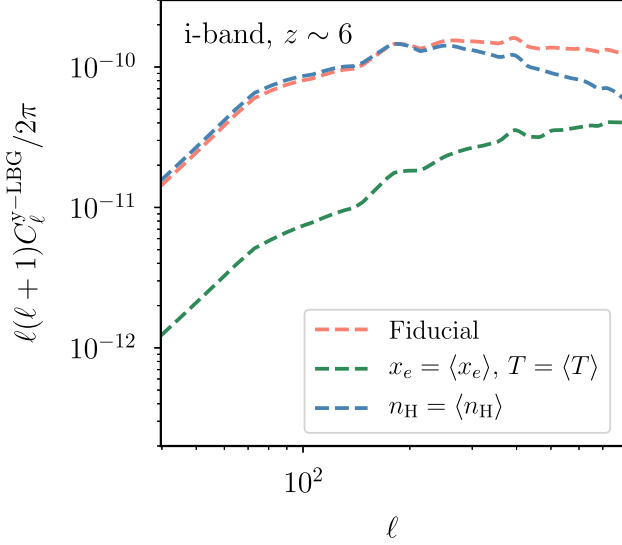
Figure 4. The impact on the large-scale LBG–y cross-power of nulling fluctuations in the density field (blue), and fluctuations in the ionization and temperature fields (green). Results are shown for the 1 Gpc simulation with galaxies corresponding to the $z \sim 6$ i-band dropout selection. Setting the ionization and temperature fields to their mean values reduces the cross-power by roughly an order of magnitude, whereas setting the density to the mean value only deviates from the fiducial case (pink curve) on smaller scales $\ell > 400$. These results suggest that the LBG–y cross-power on large scales is driven by fluctuations in the ionization and temperature fields.

When calculating correlation statistics, we use the dimensionless overdensity,

$$\delta_{\mathrm{g}}^{\mathrm{2D}} = \int_{z_{\min}}^{z_{\max}} \mathrm{d}z \, \frac{n(z)}{\bar{n}} \, \delta_{\mathrm{g}}^{\mathrm{3D}}(z), \tag{4}$$

$$\equiv \int_{z_{\min}}^{z_{\max}} \mathrm{d}z \, W_{\mathrm{g}}(z) \, \delta_{\mathrm{g}}^{\mathrm{3D}}(z), \tag{5}$$

where $n(z)$ is the redshift distribution of galaxies, and the galaxy window function is given by,

$$W_{\mathrm{g}}(z) = \frac{n(z)}{\bar{n}}. \tag{6}$$

Here, we match the window function to the desired redshift distribution of both the broad-band (Ono et al. 2018) and narrow-band (Shibuya et al. 2018) filters. In order to forecast realistic galaxy shot noise, we calibrate the luminosity selection such that we recover expected survey number densities, $\bar{n}$. For the mock LAE populations, we match to the observed densities from the SILVERRUSH survey (Ouchi et al. 2018), while for the LBG populations we match to forecast number densities from Wilson & White (2019).

2.4 Correlation statistics

Given the full 3D information content of the simulation, we first calculate the spatial power spectrum in 3D,

$$\langle \tilde{\delta}_{\mathrm{g}}(\mathbf{k_1}) \tilde{P}_{\mathrm{e}}(\mathbf{k_2}) \rangle = (2\pi)^3 \delta_{\mathrm{D}}(\mathbf{k_1} + \mathbf{k_2}) P_{\mathrm{g}P_{\mathrm{e}}}(k), \tag{7}$$

where a tilde indicates the Fourier transformed field, and $P_{\mathrm{g}P_{\mathrm{e}}}(k)$ is the cross-power spectrum between LAEs and the electron pressure. Using the Limber approximation (Limber 1953; Peebles 1980), this

can be projected to the 2D cross-power spectrum,

$$
\begin{aligned}
C_{\ell}^{gy} = \frac{\sigma_T}{m_{\mathrm{e}} c^2} & \int \mathrm{d}z \, (1+z)^2 \, \frac{n(z)}{\bar{n}} \, \frac{1}{\chi^2(z)} \\
& \times P_{\mathrm{g}P_{\mathrm{e}}} \left(k = \frac{\ell + 1/2}{\chi(z)}; z \right),
\end{aligned} \tag{8}
$$

where we have further assumed that the comoving angular diameter distance $d_{\mathrm{A}} = \chi$, valid in the limit of a flat universe. We construct the autopower spectra similarly.

3 FORECASTING METHODOLOGY

We now forecast the ability of CMB and galaxy surveys to measure C_{ℓ}^{gy}. Measurement of C_{ℓ}^{gy} must contend with three sources of noise: (1) shot noise from the galaxy survey, (2) instrument and foreground noise in the estimated Compton-y maps, and (3) the y contributions from structure at low redshift (i.e. after reionization) that is uncorrelated with the signal at high redshift. For future surveys, the y noise contribution from low-redshift structure can be comparable to or greater than that of instrument and foreground noise. However, as we discuss in more detail below, the contribution from low-redshift structure can be suppressed by masking detected haloes at low redshift.

The covariance of the C_{ℓ}^{gy} cross-spectrum in principle includes four-point combinations of the galaxy and Compton-y fields. However, because the Compton-y map is dominated by low-z y and noise – which are both uncorrelated with the reionization signal – we can ignore the connected part of the four-point function and consider only the Gaussian limit. In this limit, the error on the measured C_{ℓ}^{gy} is given by

$$
\sigma^2(C_{\ell_i}^{gy}) = \frac{1}{(2\ell + 1)\Delta\ell \, f_{\mathrm{sky}}} \left[\left(C_{\ell}^{yy-\mathrm{lowz}} + N_{\ell}^{yy} \right) \left(C_{\ell}^{gg} + \frac{1}{\bar{n}} \right) + \left(C_{\ell}^{gy} \right)^2 \right], \tag{9}
$$

where ℓ_i labels the ℓ bin, $\Delta\ell$ is the width of the bins, f_{sky} is the observed sky fraction, $C_{\ell}^{yy-\mathrm{lowz}}$ is the low-redshift contribution to the y autospectrum, N_{ℓ}^{yy} is the noise and foreground contribution to the y autospectrum, and C_{ℓ}^{gg} is the galaxy autospectrum. We ignore the contribution to the total y autopower from high-redshift y, since this contribution is subdominant compared to the instrumental noise and low-redshift contributions. Peng Oh, Cooray & Kamionkowski (2003) found the y autospectrum sourced by reionization to be roughly an order of magnitude below the low-redshift contribution using a model for which the mean y contributed by reionization is $\langle y_{\mathrm{reion}} \rangle \sim 3.6 \times 10^{-6}$. A more recent calculation by Hill et al. (2015) suggests $\langle y_{\mathrm{reion}} \rangle \sim 10^{-7}$, implying that the y autospectrum from reionization is three to four orders of magnitude lower than that from low-redshift structure (see also Namikawa et al., in preparation).

We compute the C_{ℓ}^{gg} and C_{ℓ}^{gy} terms directly from the simulations, as described in Section 2.4. We discuss the computation of $C_{\ell}^{yy-\mathrm{lowz}}$ and N_{ℓ}^{yy} in the following section.

3.1 Noise in Compton-y maps

We compute the low-redshift (i.e. post-reionization) contribution to the y autopower, $C_{\ell}^{yy-\mathrm{lowz}}$, using the halo model formalism. Details of the halo model calculation for the y autopower can be found in e.g. Vikram, Lidz & Jain (2017), Hill et al. (2018), and Pandey, Baxter & Hill (2020). We exactly follow the methodology described in Pandey
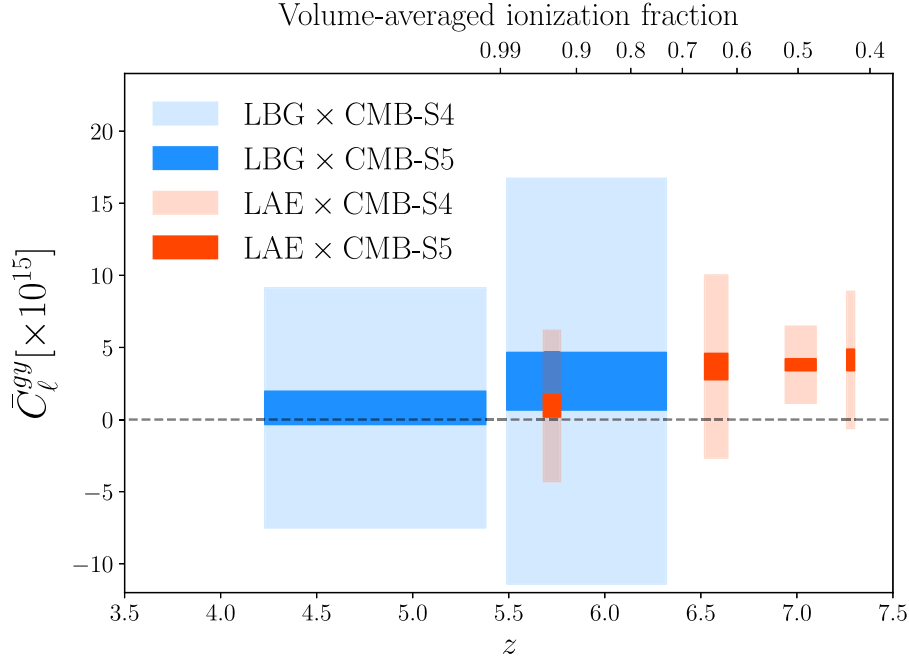
Figure 5. Forecast constraints on the amplitude of the C_ℓ^{gy} correlation as a function of z for LAEs (red) and LBGs (blue). The y-axis shows the average C_ℓ^{gy} over the range $\ell = 300$–1000 ($\ell = 300$–3000) for LBGs (LAEs), where most of the reionization signal is located. The horizontal extent of each band indicates the redshift range of the corresponding galaxy selection, while the vertical extent represents the 1σ uncertainty on the cross-power amplitude. The fainter shaded regions correspond to forecasts for y-maps from CMB-S4, while the darker regions correspond to a more futuristic survey with 10 times lower y noise than CMB-S4, i.e. what we call CMB-S5. We have assumed that haloes with $M > 3 \times 10^{13}\,\mathrm{M_\odot}$ have been identified and masked in the y-maps to reduce noise from low-redshift structure. The top x-axis shows the volume-averaged ionization fraction at the redshifts indicated by the bottom x-axis.

et al. (2020). In this model, all gas in the Universe resides in haloes. The y autospectrum can then be computed when the pressure profiles of haloes of a given mass and redshift are known, and when the clustering of the haloes is specified. We assume the halo pressure profiles are specified by the fits to hydrodynamical simulations from Battaglia et al. (2012). In these fits, the halo pressure profile is described by a generalized Navarro–Frenk–White profile (Navarro, Frenk & White 1996),

$$P(x = r/R_{200c}) = P_0(x/x_c)^\gamma (1 + (x/x_c)^\lambda)^{-\beta}, \qquad (10)$$

where r is the distance to the halo centre, R_{200c} is the halo radius, and x_0, β, λ, and γ are model parameters varied in the fits. The clustering of haloes is described by a linear bias model. We show the computed $C_\ell^{yy-\mathrm{lowz}}$ curve as the green dotted line in Fig. 2.

The integrated Compton-y signal from a halo of mass M scales roughly as $M^{5/3}$; as a result, the most massive haloes make a large contribution to the total y power despite their low abundance. Fig. 3 from Makiya et al. (2018) nicely illustrates that the bulk of the contributions to the y autospectrum come from haloes with $M \sim 10^{15}\,\mathrm{M_\odot}$. Consequently, if massive haloes can be detected and masked, the total $C_\ell^{yy-\mathrm{lowz}}$ can be reduced significantly. In Fig. 2, we show the impact on $C_\ell^{yy-\mathrm{lowz}}$ when haloes with $M > M_{\mathrm{threshold}} = 3 \times 10^{13}\,\mathrm{M_\odot}$ are removed. This mass scale corresponds roughly to large galaxy groups, which will be detected over large patches of the sky by e.g. LSST (Ivezić et al. 2019). The resultant reduction in power is nearly two orders of magnitude for $\ell \lesssim 5000$.

Although masking low-redshift haloes will not bias the high-redshift correlation measurements (since the low-redshift haloes are uncorrelated with those at high redshift), it will reduce available sky area and thereby degrade the signal to noise. However, even for the aggressive masking of haloes with $M > M_{\mathrm{threshold}} = 3 \times 10^{13}\,\mathrm{M_\odot}$,

the reduction in available sky area is only about 10 per cent if one assumes that haloes are masked with a disc of radius 1 arcmin (for the most massive haloes, this level of masking will not be sufficient, but the number of such haloes is small). One can significantly reduce the masked fraction at the cost of slightly enhanced low-redshift y noise power by increasing $M_{\mathrm{threshold}}$ by a small amount. A future analysis with actual data could determine an optimal $M_{\mathrm{threshold}}$ from a signal-to-noise standpoint. For simplicity, we will ignore the reduction in f_{sky} caused by masking below.

As we will show below, the expected signal-to-noise for the galaxy–y correlation measurements is low, even for futuristic surveys. We therefore make forecasts for both CMB Stage 4 (CMB-S4; Abazajian et al. 2016), and for a more futuristic survey (which we refer to as CMB-S5) that has a factor of 10 lower noise in Compton-y. The y noise forecasts for CMB-S4 are taken from Abazajian et al. (2019). Since the noise in the estimate of y is very roughly proportional to the temperature noise, our estimated y noise for CMB-S5 is about a factor of four better than that expected for Probe of Inflation and Cosmic Origins (PICOs; Sutin et al. 2018). While adopting such a low noise level is clearly optimistic, taking this route makes it clear how much improvement relative to CMB-S4 is needed in order to obtain interesting constraints on reionization from these measurements. The fiducial noise spectrum (for CMB-S5) is shown as the red solid curve in Fig. 2. We ignore potential issues of contamination in the y maps [such as from the cosmic infrared background (CIB)], but return briefly to this issue in Section 5.

We will adopt $f_{\mathrm{sky}} = 0.5$ for all forecasts. This is a reasonable assumption if one considers CMB-S4 or a space-based CMB survey (like PICO) correlated with galaxies identified from a wide-field survey like LSST. Dropout techniques should make selection of LBGs at high-redshift from such wide-area surveys feasible. Using

$f_{sky} = 0.5$ is much more optimistic for the LAE-based forecasts, considering that current LAE surveys cover at most $\sim 20\,deg^2$ (Ono et al. 2018). Expanded sky coverage is expected in the future from missions such as SPHEREx (Doré et al. 2014), which will enable intensity mapping of LAEs over its roughly $200\,deg^2$ deep fields. At the very least, our use of $f_{sky} = 0.5$ makes it easier to compare the LBG and LAE forecasts. Metal-line intensity mapping, such as [C II] (Gong et al. 2012) measured by the CONCERTO survey (Lagache, Cousin & Chatzikos 2018), may also be a useful tracer of high-redshift large-scale structure (Dumitru et al. 2019). We expect the *y*–[C II] cross-correlation to look similar to *y*–LBG, however we leave survey forecasts of that signal to future work.

4 RESULTS

The galaxy–*y* and galaxy–galaxy spectra measured from the simulations corresponding to Fig. 1 are shown in the top panel of Fig. 3, with the left-hand (right-hand) panels corresponding to the results for LBGs (LAEs). As apparent already at the map level in Fig. 1, the LAEs and LBGs are correlated with Compton-*y*. This is expected, as the pressure of the gas and the ionization field – which together yield the Compton-*y* field – are correlated with large-scale structure, as are the LBGs/LAEs.

Redshift evolution in the galaxy–*y* cross-spectra is also apparent in Fig. 3. At high redshift ($z \gtrsim 6$), the cross-spectra show an excess of power relative to the cross-spectra at lower redshift. This excess is due to additional power in the Compton-*y* field due to the presence of ionized bubbles at high redshift. We provide a rough calculation of the expected excess cross-power in Appendix B. As seen most clearly in the top right panel of Fig. 3, the bubbles grow larger with decreasing redshift, causing a shift in the peak of the *y*–galaxy cross-spectra to larger scales (lower ℓ). Between $z \sim 7.4$ to $z \sim 6.6$ the position of the high-ℓ falloff in power moves from roughly $\ell \sim 5000$ to $\ell \sim 2000$. The cosmological scale factor changes by only a factor of ~ 10 per cent over this period, indicating that the bubbles are indeed growing in physical size.

To determine whether the variations in the pressure or the ionization fields are driving the signal seen in Fig. 3, in Fig. 4 we show the impact of replacing the ionization field and temperature field, or the density field, with constant values. For simplicity, we plot only the cross-spectra with the LBG *i*-band sample. As seen in the figure, setting the density field to its mean value (blue dashed curve) has little impact on the cross-power at $\ell \sim 200-300$, where the feature seen in the left top panel of Fig. 3 peaks. On the other hand, setting the ionization and temperature fields to their mean values has a large impact on the measured cross-power. This suggests that the large-scale power seen in the C_ℓ^{gy} cross-spectrum is dominated by fluctuations in the ionization and temperature fields. Similar results hold for the cross-spectra with the LAEs, albeit on smaller scales at high redshift.

The bottom panels of Fig. 3 show the galaxy autopower spectra (thick curves) including the shot noise contribution (thin curves). For the most part, over the range of scales of interest for probing reionization, the galaxy clustering power dominates over the shot noise. Consequently, increasing the galaxy density is not expected to result in significant increases in the signal-to-noise of the galaxy–*y* correlations. We note that the galaxy autopower spectra are dependent on the survey selection. Changing the selection will change the galaxy bias, as well as the level of shot noise. The choice of LAE luminosity and equivalent width cut-offs for the different narrow-band surveys results in the different shot noise levels seen in the lower right panel

of Fig. 3 (see table 1 of Weinberger et al. 2020 for further details of the LAE selection).

In Fig. 5, we show the forecast errorbars on the amplitude of the *y*–LBG cross-power spectra (for both LAEs and LBGs) as a function of redshift. We have defined this amplitude as the mean value of C_ℓ^{gy} over the range of $\ell \in [300, 1000]$ for LBGs and $\ell \in [300, 3000]$ for LAEs, where most of the reionization signal is located. For the fiducial LAE forecast (solid red), we project that this amplitude can be measured with signal-to-noise of roughly 1.2σ, 4.0σ, 9.2σ, and 5.6σ in order of increasing redshift bin; the LBG correlation (solid blue) can be measured with signal-to-noise of roughly 1σ and 1.5σ in order of increasing redshift bin. The width of each band in Fig. 5 indicates the redshift extent of that particular galaxy sample.

There is a clear evolution in the correlation function amplitude shown in Fig. 5 with redshift, corresponding to the changing statistical properties of the *y* field over the course of reionization. For reference, the top *x*-axis in Fig. 5 shows the evolution of the volume-averaged ionization fraction. We caution that our fiducial forecasts have assumed a very high sensitivity CMB experiment, a factor of 10 times better than the planned CMB-S4 sensitivity. Furthermore, our assumption of $f_{sky} = 0.5$ for the LAE sample may be unrealistic. The light red and blue bands shown in Fig. 5 correspond to forecasts for CMB-S4 noise levels; in this case, the detection significance for the different cross-spectra is always below 2σ.

5 DISCUSSION

We have shown that the correlation between high-redshift LAEs and LBGs with Compton-*y* carries information about the reionization history of the Universe. The galaxy–*y* cross-spectra shown in Fig. 3 reveal distinctive features at the scales of reionized bubbles, and show clear evidence for redshift evolution over the duration of reionization. Similarly, Fig. 5 shows the redshift evolution of the average galaxy–*y* power (over a restricted ℓ range), which roughly tracks the reionization history of the Universe.

Since the galaxy power is not expected to be shot-noise limited over the relevant angular scales ($\ell \lesssim 2000$, see Fig. 3), it is noise in the Compton-*y* map that largely determines our ability to measure the LBG/LAE-*y* cross-spectrum. Both instrumental noise and *y* from low-*z* haloes are important sources of noise for our analysis. Masking haloes with $M > 3 \times 10^{13}\,M_\odot$ reduces the low-*z* *y* power by roughly two orders of magnitude, as seen in Fig. 2. In principle, less massive haloes could be masked in order to further reduce $C_\ell^{yy-low z}$ and increase the signal-to-noise of the cross-correlation measurements. However, as seen in Fig. 2, masking haloes much less massive than $3 \times 10^{13}\,M_\odot$ will cause the instrumental and foreground noise to dominate the total noise power, and further halo masking will therefore be ineffective.

Unfortunately, the amplitude of the *y*–LBG/LAE cross-spectrum is small compared to current and future noise levels in Compton-*y* maps. For expected CMB-S4 noise levels, no significant detection of the *y*–LBG/LAE cross-spectrum is expected. A futuristic survey with 10 times the sensitivity of CMB-S4, or roughly four times the sensitivity of PICO, could make a moderate-significance detection of the redshift evolution of the LAE–*y* correlation, as shown in Fig. 5, assuming a wide-field LAE sample. By the time any such CMB mission is completed, however, we will likely have tighter constraints on reionization from 21 cm observations and from measurements of the kSZ effect.

Finally, we comment that our analysis has ignored potential systematics that might be important in very low-noise Compton-*y* maps, such as contamination of *y* maps by e.g. the CIB. Since the

CIB is sourced primarily from $z \lesssim 4$ (e.g. Schmidt et al. 2014), and because the morphology of any contaminating CIB signal is likely very different from the y signal generated by the ionized bubbles, CIB contamination is unlikely to bias the C_ℓ^{gy} measurements considered here, but it could act as a significant source of noise. However, given our conclusion that even very futuristic surveys have little chance of detecting C_ℓ^{gy} at high significance, we postpone a more careful consideration of such systematics to future work.

DATA AVAILABILITY

The data used to generate the figures in this work are available upon request.

REFERENCES

Abazajian K. N. et al., 2016, CMB-S4 Science Book, 1st edn, preprint (arXiv:1610.02743)
Abazajian K. et al., 2019, preprint (arXiv:1907.04473)
Aubert D., Teyssier R., 2008, MNRAS, 387, 295
Aubert D., Teyssier R., 2010, ApJ, 724, 244
Battaglia N., Bond J. R., Pfrommer C., Sievers J. L., 2012, ApJ, 758, 75
Behnel S., Bradshaw R., Citro C., Dalcin L., Seljebotn D. S., Smith K., 2011, Comput. Sci. Eng., 13, 31
Birkinshaw M., 1999, Phys. Rep., 310, 97
Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897
Bouwens R. J. et al., 2015, ApJ, 803, 34
Choudhury T. R., Puchwein E., Haehnelt M. G., Bolton J. S., 2015, MNRAS, 452, 261
Dijkstra M., Wyithe J. S. B., 2012, MNRAS, 419, 3181
Doré O. et al., 2014, preprint (arXiv:1412.4872)
Dumitru S., Kulkarni G., Lagache G., Haehnelt M. G., 2019, MNRAS, 485, 3486
Dunlop J. S., 2013, Observing the First Galaxies, Springer, Berlin, Heidelberg, p. 223
Gong Y., Cooray A., Silva M., Santos M. G., Bock J., Bradford C. M., Zemcov M., 2012, ApJ, 745, 49
Gruzinov A., Hu W., 1998, ApJ, 508, 435
Hill J. C., Battaglia N., Chluba J., Ferraro S., Schaan E., Spergel D. N., 2015, Phys. Rev. Lett., 115, 261301
Hill J. C., Baxter E. J., Lidz A., Greco J. P., Jain B., 2018, Phys. Rev. D, 97, 083501
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Ivezić v. et al., 2019, ApJ, 873, 111
Jones E. et al., 2001, SciPy: Open source scientific tools for Python, Available at: http://www.scipy.org/
Keating L. C., Weinberger L. H., Kulkarni G., Haehnelt M. G., Chardin J., Aubert D., 2020, MNRAS, 491, 1736
Kubota K., Yoshiura S., Takahashi K., Hasegawa K., Yajima H., Ouchi M., Pindor B., Webster R. L., 2018, MNRAS, 479, 2754
Kulkarni G., Keating L. C., Haehnelt M. G., Bosman S. E. I., Puchwein E., Chardin J., Aubert D., 2019, MNRAS, 485, L24
Lagache G., Cousin M., Chatzikos M., 2018, A&A, 609, A130
Laureijs R. et al., 2011, preprint (arXiv:1110.3193)
Limber D. N., 1953, ApJ, 117, 134
Makiya R., Ando S., Komatsu E., 2018, MNRAS, 480, 3928
Matsakis N. D., Klock F. S., II, 2014, Ada Lett., 34, 103
Mesinger A., Furlanetto S., Cen R., 2011, MNRAS, 411, 955
Nasir F., D'Aloisio A., 2020, MNRAS, 494, 3080
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Ono Y. et al., 2018, PASJ, 70, S10
Ouchi M. et al., 2018, PASJ, 70, S13
Pandey S., Baxter E. J., Hill J. C., 2020, Phys. Rev. D, 101, 043525
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton
Peng Oh S., Cooray A., Kamionkowski M., 2003, MNRAS, 342, L20
Planck Collaboration XVI, 2014, A&A, 571, A16
Planck Collaboration VI, 2020, A&A, 641, A6
Puchwein E., Springel V., 2013, MNRAS, 428, 2966
Reichardt C. L., 2016, Observing the Epoch of Reionization with the Cosmic Microwave Background. Springer International Publishing, Switzerland, p. 227
Reichardt C. L. et al., 2020, preprint (arXiv:2002.06197)
Roy A., Kulkarni G., Meerburg P. D., Challinor A., Baccigalupi C., Lapi A., Haehnelt M. G., 2021, J. Cosmol. Astropart. Phys., 2021, 003
Schmidt S. J., Ménard B., Scranton R., Morrison C. B., Rahman M., Hopkins A. M., 2014, MNRAS, 446, 2696
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Shibuya T. et al., 2018, PASJ, 70, S15
Sobacchi E., Mesinger A., Greig B., 2016, MNRAS, 459, 2741
Spergel D. et al., 2015, preprint (arXiv:1503.03757)
Springel V., 2005, MNRAS, 364, 1105
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Sunyaev R. A., Zeldovich Y. B., 1972, Comments Astrophys. Space Phys., 4, 173
Sunyaev R. A., Zeldovich I. B., 1980, MNRAS, 190, 413
Sutin B. et al., 2018, Proc. SPIE, Space Telescopes and Instrumentation 2018: Optical, Infrared, and Millimeter Wave, SPIE Conf. Ser. Vol. 10698, p. 106984F
Trenti M., Stiavelli M., Bouwens R. J., Oesch P., Shull J. M., Illingworth G. D., Bradley L. D., Carollo C. M., 2010, ApJ, 714, L202
van der Walt S., Colbert S. C., Varoquaux G., 2011, Comput. Sci. Eng., 13, 22
Van Rossum G., Drake F. L., Jr, 1995, Python Tutorial. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands
Viel M., Haehnelt M. G., Springel V., 2004, MNRAS, 354, 684
Vikram V., Lidz A., Jain B., 2017, MNRAS, 467, 2315
Weinberger L. H., Kulkarni G., Haehnelt M. G., Choudhury T. R., Puchwein E., 2018, MNRAS, 479, 2564
Weinberger L. H., Haehnelt M. G., Kulkarni G., 2019, MNRAS, 485, 1350
Weinberger L. H., Kulkarni G., Haehnelt M. G., 2020, MNRAS, 494, 703
Wilson M. J., White M., 2019, J. Cosmol. Astropart. Phys., 2019, 015

APPENDIX A: THE HOT HALO GAS CONTRIBUTION AND THE EFFECT OF FEEDBACK

In this work, we are chiefly concerned with the correlations between high-redshift galaxy populations and the y-signal arising from the IGM during reionization. As discussed in Section 3.1, a key component in the low-redshift y-signal comes not from the IGM but from the hot gas within haloes (i.e. from clusters) where stellar feedback is important. In this appendix, we explore to what extent hot halo gas contributes to the high-redshift y-signal, and also whether the simulation of feedback impacts our results.

To test the contribution of hot halo gas, we recalculate the y-signal with an upper temperature limit of $10^{5.5}$ K, setting all gas above this temperature to this cap. In the left-hand panel of Fig. A1, we show a comparison using our fiducial $L = 160 \, \mathrm{cMpc} \, h^{-1}$ simulation of the effect of introducing this temperature cap. In this panel, we compare the y–LBG cross-power spectra, and find that introducing a temperature cap does not significantly affect our results. This suggests that the hot halo gas is not a dominant contribution to the large-scale correlation signal, but that as discussed in Appendix B, we expect the signal to arise predominantly from the ionized bubble distribution.

Regarding feedback, the Sherwood simulations we employ in this work were designed to capture the evolution of the low-density IGM, and approximate star formation within haloes using the QUICK_LYA feedback prescription (see Section 2; Viel et al. 2004). To test the effect of feedback, we employ two further simulations from the Sherwood suite with box sizes $L = 40 \, \mathrm{cMpc} \, h^{-1}$ and particle numbers $N = 2 \times 1024^3$: (i) a fiducial simulation run using QUICK_LYA feedback prescription (referred to as L40 N1024); (ii) a second simulation run with a more accurate feedback prescription based on the supernova-driven wind model of Puchwein & Springel (2013) (referred to as L40 N1024 ps13). In the right-hand panel of Fig. A1, we show the angular y autopower spectra for both $L = 40 \, \mathrm{cMpc} \, h^{-1}$ simulations to compare the impact of feedback. We see that while there is some deviation on small scales $\ell > 10^4$, there is good agreement in the large-scale regime that is of central interest to this work. This suggests that our results are robust to the choice of feedback prescription.

APPENDIX B: ANALYTICAL CALCULATION

In this appendix, we provide rough analytical estimates for the expected C_ℓ^{gy} signal discussed in the main text. The y signal is given by

$$y = \frac{\sigma_T}{m_e c^2} \int \mathrm{d}l \, P_e(l), \quad (B1)$$

where l is physical distance along the line of sight and P_e is the electron pressure. We consider galaxies located in a redshift bin that is not much wider than the size of the ionized bubbles. This assumption simplifies the calculation because it allows us to ignore the impact of multiple ionized bubbles along the line of sight. For the case of the LAEs – which are selected with narrow-band observations – this assumption is quite reasonable; for the LBGs, this assumption is likely to be less accurate. We make the approximation that all galaxies live at the centres of reionized bubbles. Finally, we assume that the ionization fraction and pressure within the bubble are constant (not an unreasonable approximation, see e.g. Weinberger et al. 2018). Making these approximations, the y along the direction to a galaxy becomes

$$y_{\mathrm{bubble}} = \frac{2\sigma_T R P_e}{m_e c^2 (1+z)}, \quad (B2)$$

where R is the bubble radius in comoving units. We can write the pressure in the bubble as

$$P_e \sim k_B (1+\delta) \Omega_b \rho_c (1+z)^3 \left(X_p(1/u) + Y_p(2/4u) \right) T, \quad (B3)$$

where $\delta \sim 4$ is the overdensity in a bubble, T is the bubble temperature, and u is the atomic mass unit. Assuming $T \sim 10^4$ K, we have

$$y_{\mathrm{bubble}} \sim 1.7 \times 10^{-11} (1+\delta)(1+z)^2 \left(\frac{R}{10 \, \mathrm{cMpc} \, h^{-1}} \right). \quad (B4)$$

At $z = 7$, and for $\delta = 4$, we find $y \sim 5.6 \times 10^{-9}$ for a bubble of radius $10 \, \mathrm{cMpc} \, h^{-1}$. This value agrees well with the LAE results seen in the Sherwood simulations (Fig. 1).

We now consider estimates for C_ℓ^{gy}. The multipole corresponding to the bubble size is roughly

$$\ell_{\mathrm{bubble}} \sim \pi/\theta_{\mathrm{bubble}} \sim \pi \chi(z)/R. \quad (B5)$$

From Fig. 1, at $z \sim 7$ we have $R \sim 10 \, \mathrm{cMpc} \, h^{-1}$, so $\ell_{\mathrm{bubble}} \sim 1900$; at $z \sim 6.6$, we have $R \sim 30 \, \mathrm{cMpc} \, h^{-1}$, so $\ell_{\mathrm{bubble}} \sim 640$. These values
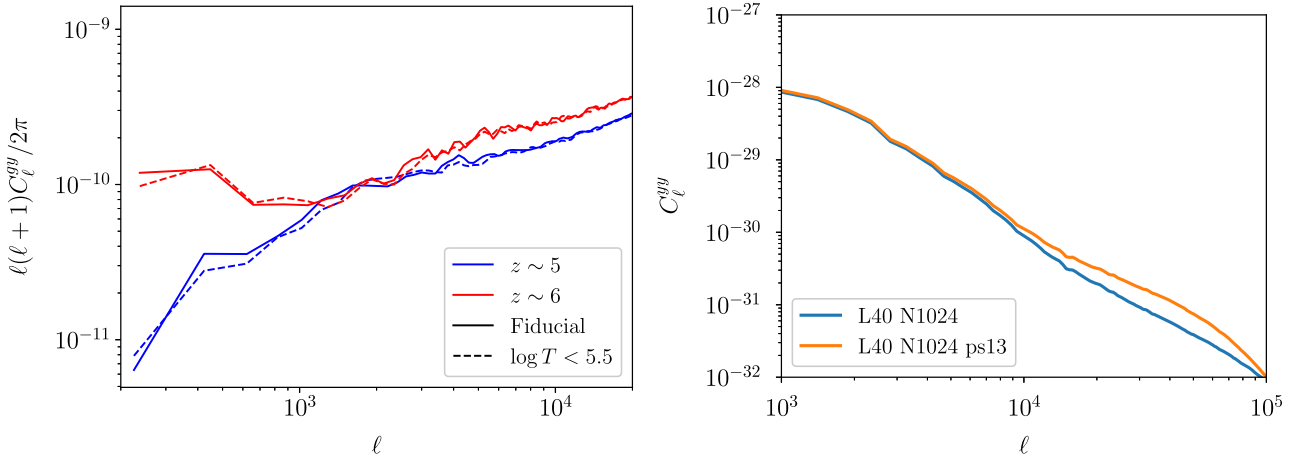
Figure A1. Testing the contribution of hot halo gas and the effects of feedback. *Left*: A comparison of the dimensionless y–LBG cross-power spectrum for our fiducial simulations (solid lines) and for the case when the temperature in the simulation is capped at $T < 10^{5.5}$ K (dashed lines). *Right*: The y autospectrum at $z = 7$ for two simulations: 'L40 N1024' (blue) that uses the fiducial feedback prescription employed elsewhere in this work, and 'L40 N1024 ps13' (orange) that uses the supernova-driven feedback model of Puchwein & Springel (2013).

agree well with the location of the upturns of $C^{gy}(\ell)$ seen in Fig. 3. This suggests that the shape of the cross-spectra directly encodes information about the bubble sizes.

The cross-power is related to the *y*-profile of the bubble via a Hankel transform:

$$C_\ell^{gy} = 2\pi \int d\theta \, \theta \, J_0(\ell\theta)\xi^{gy}(\theta), \qquad (B6)$$

where $\xi^{gy}(\theta)$ is the bubble profile as a function of angular separation, θ. Still assuming that all galaxies live at the centers of equally sized bubbles, and that the *y* profile inside a bubble is constant, we have

$$C_\ell^{gy} \sim 2\pi \int_0^{\theta_{\rm bubble}} d\theta \, \theta \, J_0(\ell\theta)y_{\rm bubble}, \qquad (B7)$$

where $\theta_{\rm bubble}$ is the angular scale corresponding to the radius of the bubble at its redshift. Performing this integral for the bubble radius

and *y* amplitude estimated above at $z \sim 7$, the predicted amplitude of the cross-power is $C^{gy}(\ell = 2000) \sim 8 \times 10^{-15}$.

Some bubbles will of course be much smaller than $10\,{\rm cMpc}\,h^{-1}$. For these bubbles, the *y* amplitude at R will be zero, and the ξ^{gy} correlation at R will therefore be suppressed. If we assume that a fraction β of galaxies live in bubbles smaller than $10\,{\rm cMpc}\,h^{-1}$, then the correlation function at this scale will be suppressed by $(1 - \beta)$. The true measured cross-power is roughly $C^{gy}(\ell = 2000) \sim 2 \times 10^{-15}$. This suggests that $(1 - \beta) \sim 0.25$. From Fig. 1, at $z \sim 7$ it seems reasonable that roughly 75 per cent of galaxies live in bubbles smaller than $10\,{\rm cMpc}\,h^{-1}$. Our analytic estimates of the cross-power and the simulation results therefore appear roughly consistent.

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