

4-25-2005

The Oxford-Dartmouth Thirty Degree Survey - II. Clustering of Bright Lyman Break Galaxies: Strong Luminosity-Dependent Bias at $z = 4$

Paul D. Allen
University of Oxford

Leonidas A. Moustakas
University of Oxford

Gavin Dalton
University of Oxford

Emily MacDonald
University of Oxford

Chris Blake
University of Oxford

See next page for additional authors

Follow this and additional works at: <https://digitalcommons.dartmouth.edu/facoa>

 Part of the [External Galaxies Commons](#)

Recommended Citation

Allen, Paul D.; Moustakas, Leonidas A.; Dalton, Gavin; MacDonald, Emily; Blake, Chris; Clewley, Lee; Heymans, Catherine; and Wegner, Gary, "The Oxford-Dartmouth Thirty Degree Survey - II. Clustering of Bright Lyman Break Galaxies: Strong Luminosity-Dependent Bias at $z = 4$ " (2005). *Open Dartmouth: Faculty Open Access Articles*. 1861.
<https://digitalcommons.dartmouth.edu/facoa/1861>

This Article is brought to you for free and open access by Dartmouth Digital Commons. It has been accepted for inclusion in Open Dartmouth: Faculty Open Access Articles by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Authors

Paul D. Allen, Leonidas A. Moustakas, Gavin Dalton, Emily MacDonald, Chris Blake, Lee Clewley, Catherine Heymans, and Gary Wegner

The Oxford–Dartmouth Thirty Degree Survey – II. Clustering of bright Lyman break galaxies: strong luminosity-dependent bias at $z = 4$

Paul D. Allen,^{1,2*} Leonidas A. Moustakas,^{1,3} Gavin Dalton,^{1,4} Emily MacDonald,¹ Chris Blake,^{1,5} Lee Clewley,¹ Catherine Heymans^{1,6} and Gary Wegner⁷

¹*Astrophysics, University of Oxford, Keble Road, Oxford OX1 3RH*

²*RSAA, The Australian National University, Mount Stromlo Observatory, Cotter Rd, Weston, ACT 2611, Australia*

³*Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA*

⁴*Space Science and Technology Division, Rutherford Appleton Laboratory, Didcot OX11 0QX*

⁵*School of Physics, University of New South Wales, Sydney, NSW 2052, Australia*

⁶*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117, Heidelberg, Germany*

⁷*Department of Physics and Astronomy, Dartmouth College, 6127 Wilder Laboratory, Hanover, NH 03755, USA*

Accepted 2005 April 25. Received 2005 April 22; in original form 2003 August 11

ABSTRACT

We present measurements of the clustering properties of bright ($L > L_*$) $z \sim 4$ Lyman break galaxies (LBGs) selected from the Oxford–Dartmouth Thirty Degree Survey (ODT). We describe techniques used to select and evaluate our candidates and calculate the angular correlation function, which we find best fitted by a power law, $\omega(\theta) = A_w \theta^{-\beta}$ with $A_w = 15.4$ (with θ in arcsec), using a constrained slope of $\beta = 0.8$. Using a redshift distribution consistent with photometric models, we deproject this correlation function and find a comoving $r_0 = 11.4_{-1.9}^{+1.7} h_{100}^{-1}$ Mpc in a $\Omega_m = 0.3$ flat Λ cosmology for $i_{AB} \leq 24.5$. This corresponds to a linear bias value of $b = 8.1_{-2.6}^{+2.0}$ (assuming $\sigma_8 = 0.9$). These data show a significantly larger r_0 and b than previous studies at $z \sim 4$. We interpret this as evidence that the brightest LBGs have a larger bias than fainter ones, indicating a strong luminosity dependence for the measured bias of an LBG sample. Comparing this against recent results in the literature at fainter (sub L_*) limiting magnitudes, and with simple models describing the relationship between LBGs and dark matter haloes, we discuss the implications on the implied environments and nature of LBGs. It seems that the brightest LBGs (in contrast with the majority sub L_* population) have clustering properties, and host dark matter halo masses, which are consistent with them being progenitors of the most massive galaxies today.

Key words: surveys – galaxies: evolution – galaxies: fundamental parameters – galaxies: high-redshift – galaxies: statistics.

1 INTRODUCTION

The study of the Universe at very high redshifts has expanded rapidly over the last decade. It is now possible to observe galaxies over more than 90 per cent of the age of the Universe. One of the most significant breakthroughs has been the discovery of a population of strongly clustered, star-forming galaxies at $2.5 < z < 4.5$, using the Lyman break technique pioneered by Steidel & Hamilton (1992). It is possible to select significant numbers of these galaxies using deep ground-based multicolour imaging (e.g. Steidel et al. 1996).

Lyman break galaxies (LBGs) represent the largest known population of high-redshift objects, and therefore present a window into an important stage in the formation of galaxies and large-scale

structure. Much is now known about their physical characteristics (Pettini et al. 2001; Shapley et al. 2001, 2003). They are somewhat dusty starburst galaxies with star formation rates in the range of 10 to a few hundred $M_\odot \text{ yr}^{-1}$, contributing a highly significant fraction of the stars formed at $z \sim 2.5$ –5 (Adelberger & Steidel 2000).

Comparison of their clustering properties and number densities with semi-analytical models suggests that they are either relatively small galaxies, experiencing brief and infrequent bursts of star formation that are primarily driven by galaxy–galaxy mergers (‘the collisional starburst model’; Somerville, Primack & Faber 2001), or in very massive environments with large reservoirs of gas, becoming massive L_* galaxies today (‘the massive halo model’; Steidel et al. 1996; Baugh et al. 1998).

Although these scenarios have been tested against models of galaxy formation using the LBG angular correlation function on different scales, and as a function of luminosity (Wechsler et al.

*E-mail: paul@mso.anu.edu.au

2001; Bullock, Wechsler & Somerville 2002) the results are still somewhat dependent on the observational sample used. Correlation scale measurements range from $r_0 = 2\text{--}12 h_{100}^{-1}$ Mpc (Adelberger et al. 1998; Giavalisco et al. 1998; Steidel et al. 1998; Arnouts et al. 1999; Giavalisco & Dickinson 2001; Ouchi et al. 2001; Adelberger et al. 2003; Foucaud et al. 2003). The range is likely due in part to cosmic variance because of relatively small sample sizes and may also reflect the luminosity dependence of clustering and the effects of small-scale clustering.

The data (especially in the more studied $z \sim 3$ population) do not yet provide unequivocal answers. Porciani & Giavalisco (2002) find a lack of power in the angular correlation function on small scales, suggesting few LBG close pairs. Giavalisco & Dickinson (2001) and Foucaud et al. (2003) find fainter LBGs less strongly clustered than brighter ones. Conversely, the clustering results at $z \sim 4$ of Ouchi et al. (2001) imply an excess of LBG close pairs from which they estimate an LBG merger rate. Using $z \sim 3$ clustering data provided by Adelberger et al., several authors (Somerville et al. 2001; Wechsler et al. 2001; Bullock et al. 2002) have shown that the collisional starburst type models are marginally more favourable than massive halo models. However, it seems clear that current data sets are not extensive enough, and do not have accurate enough photometric redshifts in large samples, to provide very strong constraints on the relationship between LBGs and their host dark matter haloes. To begin to explore these questions, it is useful to have very wide and deep multiband imaging, to optimize the selection and redshift determination for many thousands of LBG candidates, over a range in absolute magnitudes. The Oxford–Dartmouth Thirty Degree Survey (ODT) is one such survey.

The structure of this paper is as follows. In Section 2 we present the ODT survey, and its salient characteristics. The candidate selection techniques, and their limitations, are discussed in Section 3. The general statistical properties of $z \sim 4$ LBGs are presented in Section 4, and the projected and spatial correlation functions are the topic of Section 5. In Section 6, these results are placed in a theoretical context, focusing on the nature of the environments in which LBGs exist, the main conclusions finally being drawn in Section 7. Throughout this paper we use AB magnitudes, and assume a $(\Omega_m, \Omega_\Lambda, \sigma_8) = (0.3, 0.7, 0.9)$ cosmology, unless otherwise stated, with $H_0 = 100 h_{100} \text{ km s}^{-1} \text{ Mpc}^{-1}$.

2 OXFORD–DARTMOUTH THIRTY DEGREE SURVEY

The ODT is a deep-wide survey using the Wide Field Camera (WFC) on the 2.5-m Isaac Newton Telescope (INT) at La Palma. The survey employs six broad-band filters in $UBVRi'Z$, and the transmission curves are provided in Fig. 1. When completed, the survey will cover $\sim 30 \text{ deg}^2$ in $BVRi'Z$ to $R_{5\sigma} = 25.25$ in four fields of $5\text{--}10 \text{ deg}^2$ each. The coordinates of these four fields are provided in Table 1. Presently, approximately 25 deg^2 of the survey have been observed in $BVRi'Z$, although only data from the Andromeda field are analysed in detail here. The U data have only been observed in the best conditions, and currently cover $\sim 1 \text{ deg}^2$, all of which is in the Andromeda field. We are also currently undertaking a K -band survey to $K_{5\sigma} \approx 18.5$ using the 1.3-m McGraw-Hill Telescope at the MDM Observatory on Kitt Peak (Olding 2002). The three largest fields are also covered at radio frequencies with the Very Large Array (VLA) in A and D array at 1.4 GHz, and the Lynx field is covered by low-frequency radio observations with the VLA in A array at 74 and 330 MHz. In addition, we are able to obtain redshifts for several sources in the survey which are part of the Texas–Oxford One-

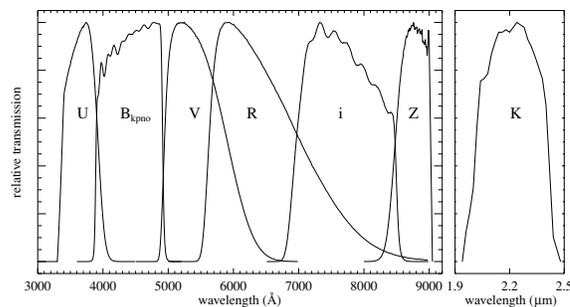


Figure 1. Filter curves for the full set of filters employed in the ODT survey, convolved with the chip QE curve.

Table 1. $3^\circ \times 3^\circ$ field centres for the ODT survey in equatorial and Galactic coordinates.

Field	α (J2000)	δ (J2000)	l	b
Andromeda ^a	00 18 24	+34 52 00	115	−27
Lynx ^{a,b}	09 09 45	+40 50 00	181	+42
Hercules ^a	16 39 30	+45 24 00	70	+41
Virgo	13 40 00	+02 30 00	330	+62

^aFields with 1.4-GHz radio data.

^bField with 74-MHz and 330-MHz radio data.

Thousand (TOOT) redshift survey of radio sources (Hill & Rawlings 2003).

Whilst the ODT survey has a number of scientific aims, one of the principal projects is the detection of large numbers of LBGs (at $z \sim 4$) over a wider field than previous studies. The ODT survey reaches depths comparable to previous studies of LBGs (e.g. Steidel et al. 1999; Ouchi et al. 2001) but over a significantly larger area of sky (cf. $\sim 900 \text{ arcmin}^2$ in Ouchi et al. 2001). The ODT therefore provides an ideal data set for the selection of a sample of the brightest LBGs.

The fields discussed in this paper were observed with the 2.5-m INT during 1998 August and 2000 September. Approximate 5σ (isophotal) depths for the data along with exposure times are shown in Table 2. The median seeing for these observations is ~ 1.1 arcsec.

The reduction of CCD frames was carried out using the IRAF package and photometry calibrated for each frame by observing fields of Landolt (1992) on photometric nights. The thinned chips in the INT

Table 2. Exposure times and detection limits (AB at 5σ isophotal) for the different pass-bands used for the ODT survey. With the exception of the U images (which use six), each pointing is broken up into three separate exposures with 5-arcsec offsets of the telescope. The detection limits are approximately equal to 5σ detections using a 2-arcsec circular aperture and 3σ detections with a 3.3-arcsec circular aperture.

Filter	Exposure time	5σ limit
U	$6 \times 1200s$	26.0
B	$3 \times 900s$	26.0
V	$3 \times 1000s$	25.5
R	$3 \times 1200s$	25.3
i'	$3 \times 1100s$	24.5
Z	$1 \times 600s$	22.5

WFC mean that fringing of sky lines becomes a problem in i' , Z and, to a lesser extent, R data. Fringing was removed from each image in turn by using a fringe frame generated by the combination of individual images for each chip. This method was successful for the i' - and R -band data, but the Z -band data have more significant fringing, and little of the acquired data have been incorporated to date. Image detections were carried out using the *SEXTRACTOR* package (Bertin & Arnouts 1996), although we chose to use our own background subtraction method, due to significant background gradients present in some of our images.

Each WFC pointing was arranged in a diagonal grid to cover each field, with several overlaps between adjacent pointings. This allows objects in overlaps to be matched and a common photometric zero-point to be applied to each field, using the method of Glazebrook et al. (1994). This method was used to obtain a common zero-point for the V -band data, which have the smallest scatter between chips and are not afflicted by fringing from sky lines.

In order to obtain consistent colours throughout the survey, photometry for other bands was corrected relative to the V -band data. This was done by minimizing the deviation of the colours of ODT stellar objects from the colours of stars from Pickles (1998) in the colour-colour plane. The ODT survey and our data reduction process are described in more detail in MacDonald et al. (2004).

3 CANDIDATE SELECTION

The LBG population of galaxies is traditionally selected by exploiting the break in their spectra shortwards of the Lyman limit ($\lambda_{\text{rest}} = 912 \text{ \AA}$). The break is caused by an intrinsic drop in the spectra of the massive stars which dominate the LBG spectrum and, more importantly, the absorption of photons shortwards of the Lyman limit by H I in the interstellar medium of the LBG. The break is further accentuated by absorption from Lyman limit systems along the line of sight to the galaxy. By choosing appropriate filter systems to straddle this break, it becomes possible to select high- z galaxy candidates.

The 'standard' method of LBG selection uses a three-band filter set to isolate high- z galaxies in the colour-colour plane (e.g. Steidel et al. 1999; Ouchi et al. 2001). The selection region is defined using the predicted colours of synthetic galaxy spectra at high redshifts; where spectroscopy is available, this can often be refined. However, for many surveys such as the ODT or the *Hubble Deep Field (HDF)*, further information from extra filters may be available. By using this extra information we can obtain photometric redshifts for objects in a survey (e.g. Arnouts et al. 1999, 2002), and select the high-redshift objects. It is also possible, with multiband data, to consider more than one filter combination, as we discuss in the next section.

3.1 Models of high- z galaxy colours

In order to define our selection criteria for LBG candidates, we plot colours for model galaxies (see Figs 2 and 3) using extended Coleman, Wu & Weedman (1980) empirical spectral templates convolved with the INT WFC filter set. We consider two different regimes: selection using $B-V/V-i'$ colours and selection with $B-R/R-i'$ colours. In addition, we plot stellar colours using the stellar spectral energy distribution (SED) library of Pickles (1998). Galaxy colours are also adjusted for absorption from the intergalactic medium using the method of Madau (1995). This attenuation of flux due to the opacity of the intergalactic medium proves to be a significant contributor to the final observed colours of high- z objects. In addition to these empirical templates, we also compare with the model grids of Stevens & Lacy (2001), which are based on the Pegase models of Fioc & Rocca-Volmerange (1997) and the

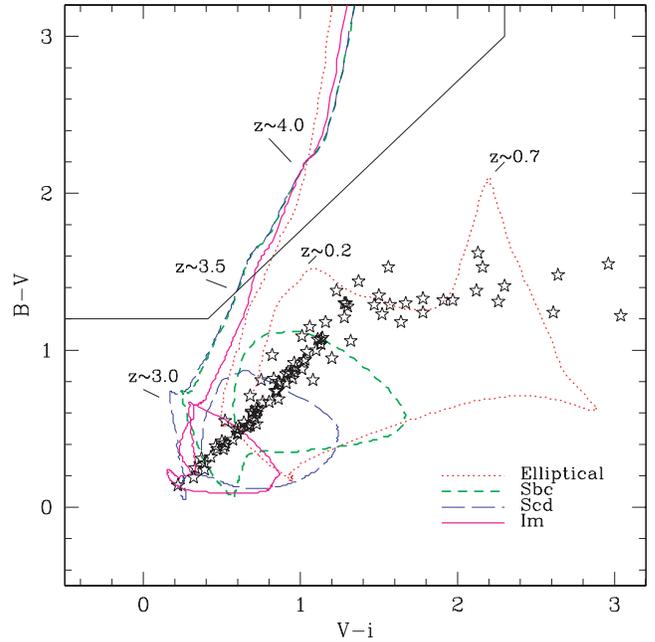


Figure 2. Model $B-V$ versus $V-i'$ colours for galaxies to $z \sim 5$. Four SEDs from Coleman et al. (1980) are used. No evolution is assumed, but the SEDs are modified to account for the effects of attenuation due to intergalactic absorbers at high redshift (Madau 1995). The four SED types of Coleman et al. (1980) are shown as a red dotted line (type E), a green short-dashed line (type Sbc), a blue long-dashed line (type Scd) and a magenta solid line (type Im). Also shown as black open stars are the colours of Galactic stars, using the stellar library of Pickles (1998). Important redshifts are labelled, most importantly, galaxies with $z > 3.5$ (i.e. LBGs) and, on the elliptical track, intermediate redshift ($z = 0.2-1.1$) objects, which are a prime source of contamination in a high-redshift sample (see Section 3.3). The final ODT selection window for selecting $z \sim 4$ objects is bounded by the solid black line.

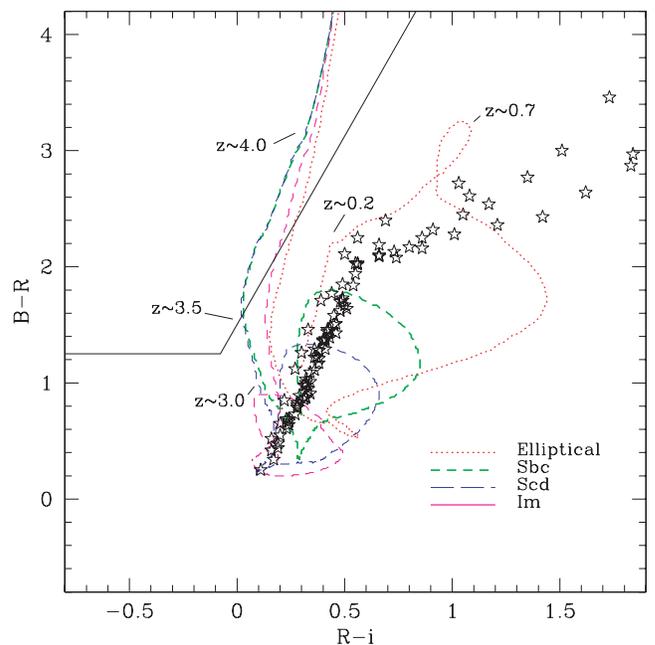


Figure 3. As Fig. 2 but using model $B-R$ versus $R-i'$ colours.

standard stars of Landolt (1992), again using our INT filter set. These models contain SEDs with a range of starburst and star formation histories along with different amounts of dust. However, they compare favourably with the simple model colours shown in Figs 2 and 3. The colours of $z \sim 4$ objects are dominated by the effects of the Lyman break, and LBGs occupy essentially the same region of colour space in both models.

In order to efficiently select high-redshift objects, we define the following colour–colour cuts based upon the model galaxy colours shown in Fig. 2 for $B-V/V-i'$ selection. For $-1.0 < (V-i') < 0.41$

$$(B-V) > 1.2. \quad (1)$$

For $0.41 < (V-i') < 2.3$

$$(B-V) > 0.95(V-i') + 0.81. \quad (2)$$

Objects with $(V-i') < -1.0$ and $(V-i') > 2.3$ are excluded. For $B-R/R-i'$ selection, we use the following colour cuts based on Fig. 3. For $-1.0 < (R-i') < -0.07$

$$(B-R) > 1.25. \quad (3)$$

For $-0.07 < (R-i') < 0.9$

$$(B-R) > 3.25(R-i') + 1.5. \quad (4)$$

Objects with $(R-i') < -1.0$ and $(R-i') > 0.9$ are excluded.

3.2 Colour–colour selection

Colours for a small subsection of the ODT survey are shown in Figs 4 and 5 along with the LBG selection region in colour–colour

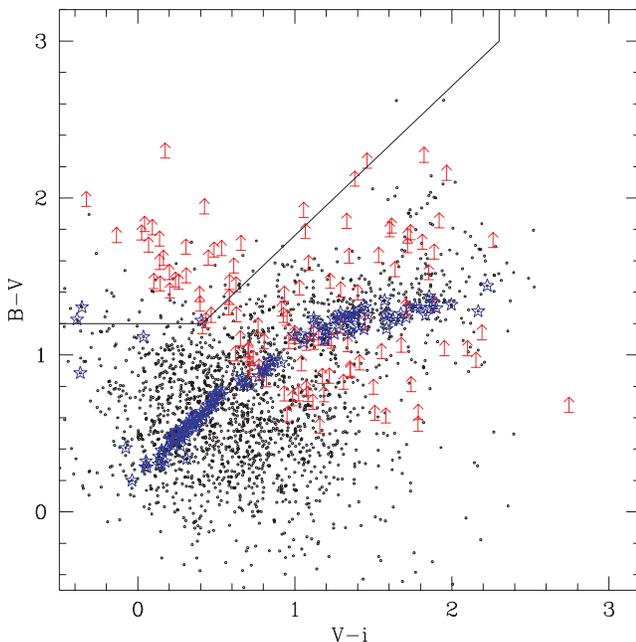


Figure 4. $B-V$ versus $V-i'$ colours for ODT data. For clarity, only 5000 randomly chosen objects are shown. The resulting colour distribution is representative of the full data set. All objects with detections in V , R and i' down to the limits of the survey (Table 2) are shown. Objects with no B detection are shown as red arrows. Stellar objects are designated by blue open stars; these were selected using SEXTRACTOR's stellarity parameter, and objects with stellarity > 0.8 and $R < 23.0$ are plotted. All other objects are shown as black circles. The stellar locus is clearly visible and this provides a useful cross-check of the photometric consistency between fields. The selection region determined from the models discussed in Section 3.1 is bounded by the black solid line.

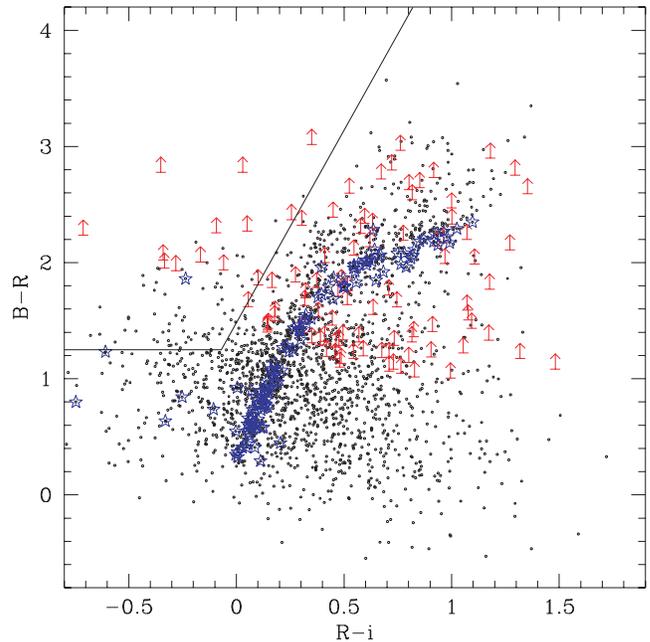


Figure 5. As Fig. 4 but using the $B-R$ versus $R-i'$ colours for a 5000 object subset from the ODT.

space. We also require a detection in the V , R and i' bands, increasing the probability of only detecting real objects close to the faint limits of the survey. In addition, only candidates fainter than $i' = 23.0$ are considered, in order to reduce contamination from lower-redshift ellipticals with similar colours to LBGs (Steidel et al. 1999). Finally, candidates that have $B > 25.5$, or are not detected by SEXTRACTOR in B , are treated as non-detections and are given a limiting magnitude of $B = 25.5$.

Given that we have four-band information available to us, we impose the additional criterion that the object lies in the selection regions of both $B-V/V-i'$ and $B-R/R-i'$ (Figs 4 and 5). Because the surface densities of objects are much higher just outside the selection region than just inside, low-redshift objects are much more likely to be scattered into the selection region than vice versa. This is somewhat reduced by requiring that an object is selected in both planes, and indeed we obtain candidate surface densities comparable to other searches for $z \sim 4$ objects (see Fig. 9). The contamination of the sample, and the effects of this restriction are discussed further in the next section.

3.3 Contamination

Although our selection criteria are fairly conservative, contamination of a non-spectroscopic sample such as this is expected. The prime candidates for contamination are foreground Galactic stars, high-redshift quasars and, most importantly, elliptical galaxies with $0.2 < z < 1.1$. Spectroscopic follow-up of a sample selected using similar selection criteria by Steidel et al. (1999) confirmed that about 20 per cent of candidates were in fact elliptical galaxies at intermediate redshifts. The peak of the luminosity function for elliptical galaxies with $z \sim 0.7$ corresponds to galaxies with $i' \sim 22.0$ (Bell et al. 2004). Having a bright cut-off of $i' = 23.0$ for the sample should reduce contamination from these objects significantly, although a number of faint potential contaminants are expected to remain. In the next section, we consider two faint limits of $i' = 24.0$

Table 3. Summary of expected contamination for stars and elliptical galaxies. The contamination fraction can be significantly reduced by using both selection criteria.

Magnitude range i'	Stars			Ellipticals			Total		
	BVi' per cent	BRi' per cent	Both per cent	BVi' per cent	BRi' per cent	Both per cent	BVi' per cent	BRi' per cent	Both per cent
23.0–24.0	22	30	8	27	4	4	49	34	12
23.0–24.5	20	28	5	15	7	3	35	35	8

and $i' = 24.5$, and so estimate contamination levels for both limits here.

In Figs 2 and 3, the elliptical track is noticeably more extended than for other types. Objects in the range $0.2 < z < 1.1$ are both close enough, and faint enough, to scatter into the LBG selection region. In order to estimate the potential level of contamination from these objects, we simulate their expected colours by combining their colours from an SED with typical photometric errors. We then normalize their numbers with an appropriate luminosity function, and test how many objects would meet our selection criteria. To obtain the colours of elliptical galaxies, we use the SED from Coleman et al. (1980), as plotted in Figs 2 and 3. The expected number of galaxies can then be obtained by integrating the galaxy luminosity function using (Gardner 1998)

$$n(m, z) dm dz = \frac{\Omega}{4\pi} \frac{dV}{dz} \phi(m, z) dm dz, \quad (5)$$

where $\phi(m, z)$ is the galaxy luminosity function and dV/dz is the comoving volume element at redshift z . We use the R -band luminosity function for elliptical galaxies from the ESO-Sculptor survey at $z \sim 0.5$ (de Lapparent et al. 2003). The rest-frame R -band and redshift range covered by this luminosity function compares favourably to the objects of interest here. Little or no evolution of the luminosity function is expected in our redshift range of interest (Bell et al. 2004). To simulate photometric errors, a random error is generated using a probability distribution describing the typical error as a function of magnitude. The error in each band is then added to the respective magnitude. If just one combination of colours is used (especially $B-V/V-i'$), then the contamination can be quite considerable (up to 27 per cent; see Table 3). However, by requiring that LBGs meet both selection criteria ($B-V/V-i'$ and $B-R/R-i'$), the expected contamination from intermediate redshift galaxies can be reduced significantly. For $i' < 24.0$, we find a contamination of 4 per cent, and for $i' < 24.5$ we obtain 3 per cent.

The ODT Andromeda field (i.e. the data discussed in this paper) has a fairly low Galactic latitude of $l = -27^\circ$, and therefore Galactic stars must be considered as potential contaminants. At the magnitudes considered here, the *SEXTRACTOR* star–galaxy classifier (in general) fails to distinguish cleanly between stellar and galactic light profiles. In particular, red stars are prone to be found in the selection window due to extreme intrinsic colours, or because they lie close to the selection window and are scattered in because of photometric errors, or a combination of both. In order to estimate the contamination due to stars, a model of the Galaxy (Robin & Creze 1986; Robin et al. 1996) is used to generate a mock catalogue¹ of Galactic stars at the Galactic latitude and longitude of the ODT Andromeda field. This model contains elements from the thin and thick discs, along with the spheroid of the Galaxy, and a simple model of Galactic extinction (Robin & Creze 1986). The V -band

number counts generated by this model were compared with the ‘Bahcall–Soneira’ model (Bahcall 1986), and with stellar number counts over the range $18 < V < 22$ from the data, and found to be consistent.

The model produces (without photometric errors) the expected mixture of stars at given Galactic coordinates in Johnson–Cousins $UBVR_C I_C$ (Vega) magnitudes. These were then transformed to the ODT $UBVRi'$ (AB) system and combined with a function to simulate a typical photometric error in each band, as done for the simulated galaxies. Our selection criteria were then applied to these stars. In total, 17 stars (in 1 deg^2) are found to match the selection criteria to $i' = 24.5$, seven of which have $i' < 24.0$. Given the surface densities of LBG candidates (see Section 4.3), this corresponds to a contamination of 8 per cent for $i' < 24.0$, and 5 per cent for $i' < 24.5$. We note that, again, these results increase significantly if only $B-V/V-i'$ or $B-R/R-i'$ selection is used (see Table 3).

Finally, using the quasar luminosity function at $z \sim 4$ (Fan et al. 2001), we expect that even if all high- z quasars were to have colours consistent with LBGs, their expected number densities make any contamination from this population negligible. Our final contamination levels are therefore 12 per cent for $i' < 24.0$ and 8 per cent for $i' < 24.5$. The contamination results are summarized in Table 3. These results compare favourably with other LBG surveys (e.g. Ouchi et al. 2001), although the overall contamination level is reduced through the advantage of having two sets of selection criteria. Objects are less likely to be scattered into both selection windows through photometric errors than just one. We also note that our observed surface densities are fully consistent with other surveys (see Section 4.3), implying similar levels of contamination and completeness.

3.4 ODT LBG samples

This paper uses the first 2 deg^2 of ODT data in the Andromeda field, which was the first part of the ODT survey to be reduced and analysed. Unfortunately, the quality and depth of the survey are variable, and we restrict the data used in this paper to those regions with the best seeing. In order to reduce contamination from any spurious signal caused by field-to-field variations, we consider two samples of data, one to $i' < 24.5$ and the other to $i' < 24.0$. These were chosen to provide a contiguous area covered by all colours to the required depth, where we are confident of consistent completeness in all bands. When fields overlap, the overlap regions are split equally (in RA) between the two fields, and objects in the region will have photometry measured from one of the fields.

The properties of the fields used (including depths) are summarized in Table 4, and Table 5 summarizes the properties of the two selected samples. All five pointings are used in the bright ($i' < 24.0$) sample. These cover a region with good seeing and a completeness beyond the required $i' < 24.0$. The distribution of these objects is shown in Fig. 6. After removal of satellite trails, overlaps, diffraction

¹ See <http://www.obs-besancon.fr/www/modele/>.

Table 4. Data for ODT Andromeda fields used in this paper. These fields form a single contiguous field. Column 1 gives the field name. Column 2 gives the RA of field centre (hms). Column 3 gives Dec. of field centre (dms). Column 4 gives seeing (averaged over $BVRi'$ bands). Column 5 gives the area in arcmin² (after removal of bad areas of the chip, bright stars, satellite trails, etc.). The final four columns give the magnitude at which the total number counts turn over in each field.

Field	RA	Dec.	Seeing	Area	B	V	R	i'
I012 ^a	00 11 38.47	+35 39 36.7	0.95	675.7	25.8	25.1	25.1	24.7
I013	00 13 49.66	+35 45 47.8	1.07	460.7	25.7	25.1	24.6	24.4
I015	00 16 02.41	+35 51 51.3	1.24	424.7	25.5	25.0	25.0	24.5
I026	00 10 29.80	+35 03 39.3	1.09	538.7	25.5	24.6	24.8	24.5
I029	00 18 19.46	+35 34 47.4	1.16	458.8	25.5	25.1	25.0	24.4

^aField used for the $i < 24.5$ LBG sample.

Table 5. Summary of ‘bright’ and ‘faint’ LBG samples selected from the ODT.

Sample	Limit (i')	Fields	Number	Total area
Bright	24.0	All	74	0.81 deg ²
Faint	24.5	I012	66	676 arcmin ²

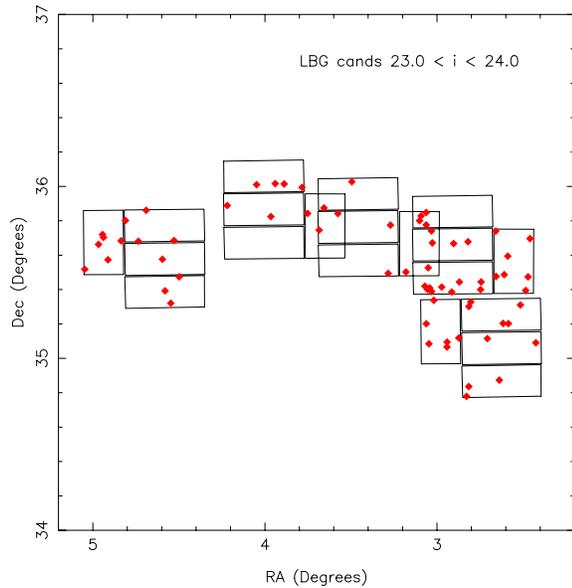


Figure 6. The distribution of our ‘bright’ sample of LBG candidates, with $23 < i' < 24$. The outlines of the chips which make up the INT wide field camera pointings for this mosaic are shown.

spikes and bright objects, the total area covered is 0.83 deg². This sample consists of 74 objects. The ‘faint’ ($i' < 24.5$) sample consists of just one INT WFC pointing covering 676 arcmin² (field I012 in Table 4). The distribution of these objects is shown in Fig. 7. This sample is made up of 66 objects.

4 GENERAL PROPERTIES OF LBGs

4.1 Redshifts

Although the primary selection criterion for LBG candidates is the effect of the integrated opacity by intergalactic H I, redshifts are an important part of the analysis and interpretation of such samples. Obtaining spectroscopic redshifts for more than a very small fraction

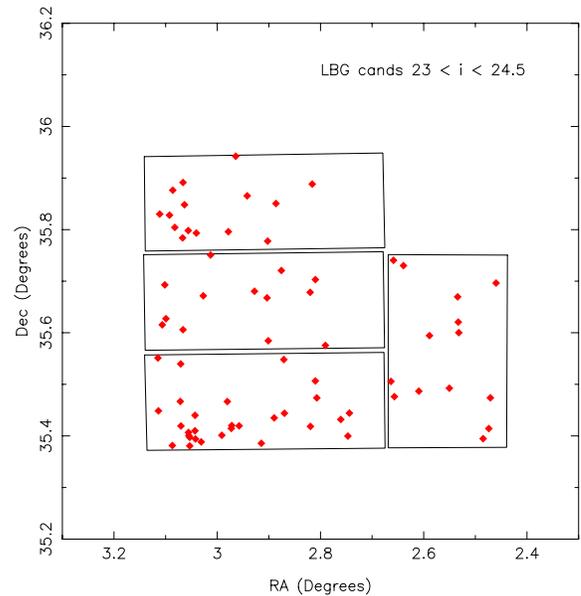


Figure 7. The distribution of our ‘faint’ sample of LBG candidates, with $23 < i' < 24.5$. The outlines of the four chips which make up the INT wide field camera are shown.

of our LBG candidates is not currently forthcoming, and represents a major challenge for current instrumentation, even on 8-m class telescopes.

However, in this paper, their most important use is in their statistical distribution, $N(z)$, and the corresponding selection function. This is used for deprojecting the angular correlation function to retrieve the spatial correlation function $\xi(r)$, and for detailed studies of the effects of different dark matter halo occupation function parametrizations on the global statistics, but not for the selection of LBG candidates.

4.2 Selection function

The selection function, as represented by the redshift distribution, is approximated by a simple analytical model based on the combined results from the colour–colour diagrams in Figs 2 and 3, and the models of Stevens & Lacy (2001). The main use of our redshift distribution will be the deprojection of the angular correlation function to obtain r_0 (Section 5), and we model redshift distributions using

$$N(z) = \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(z - \bar{z})^2}{2\sigma_z^2}\right], \quad (6)$$

where \bar{z} is the mean and σ_z is the standard deviation.

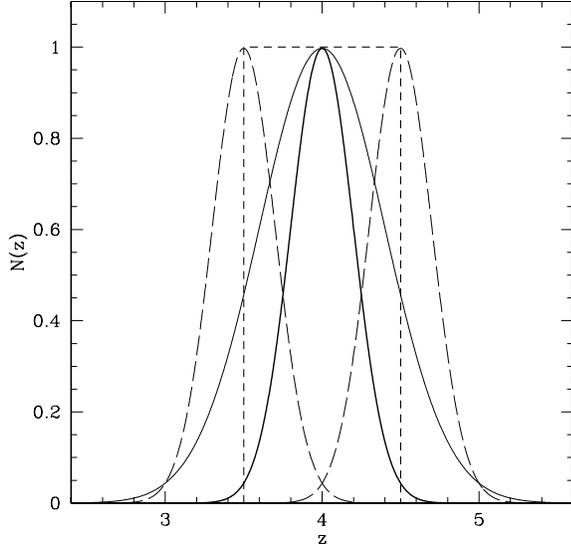


Figure 8. Normalized fiducial redshift distribution consistent with photometric models (bold solid line corresponding to a Gaussian with $\bar{z} = 4.0$ and $\sigma_z = 0.2$). The other lines show the range of selection functions considered to test the sensitivity of the Limber deprojection to the assumed redshift distribution. The bold short-dashed line shows a ‘top-hat’ function centred on $\bar{z} = 4.0$ with $\bar{z} \pm 0.5$. The thin long-dashed lines correspond to Gaussians with $\bar{z} = 3.5$ and 4.5 with $\sigma_z = 0.2$. Finally, the thin solid line corresponds to a much wider redshift distribution with $\sigma_z = 0.4$.

We find that $\bar{z} = 4.0$ and $\sigma_z = 0.2$ are a reasonable fiducial set of values, which reproduces the photometrically determined distributions. In order to test the dependence of r_0 on $N(z)$, we also consider other selection functions with varying mean redshifts, \bar{z} , and dispersions, σ_z , as well as modelling the redshift distribution as a simple top-hat function. These selection functions are shown in Fig. 8. However, as we discuss in Section 5.2, at these redshifts, the deprojected r_0 we obtain is quite insensitive to the $N(z)$ used. We do not propagate systematic errors from the uncertainty in $N(z)$ in what follows.

4.3 Surface and space densities

The surface density of objects is calculated by selecting out the regions contaminated by bright stars, chip edge-effects, satellite trails, exceptionally poor seeing and strong remnant fringe effects. The sample selection described here is performed in the i' band, requiring detection in the R and V bands as well. The measured surface densities, with Poisson uncertainties, are given in Table 6 and shown in Fig. 9. In calculating these surface densities, we only use those images that are both deep enough and have good seeing as described in Section 3.2. Fig. 9 indicates that the ODT LBG surface densities compare well to other work (Steidel et al. 1999; Ouchi et al. 2001).

Table 6. The surface and space densities for different i' -band selection magnitude ranges. Surface densities are per half magnitude bin.

Magnitude range (i')	Σ (arcmin $^{-2}$)	n_{Σ} (h^3_{100} Mpc $^{-3}$)
23.0–23.5	0.0065 ± 0.0041	$6.29 \pm 5.30 \times 10^{-6}$
23.5–24.0	0.0165 ± 0.0085	$1.60 \pm 1.10 \times 10^{-5}$
24.0–24.5	0.0685 ± 0.0101	$6.63 \pm 1.31 \times 10^{-5}$

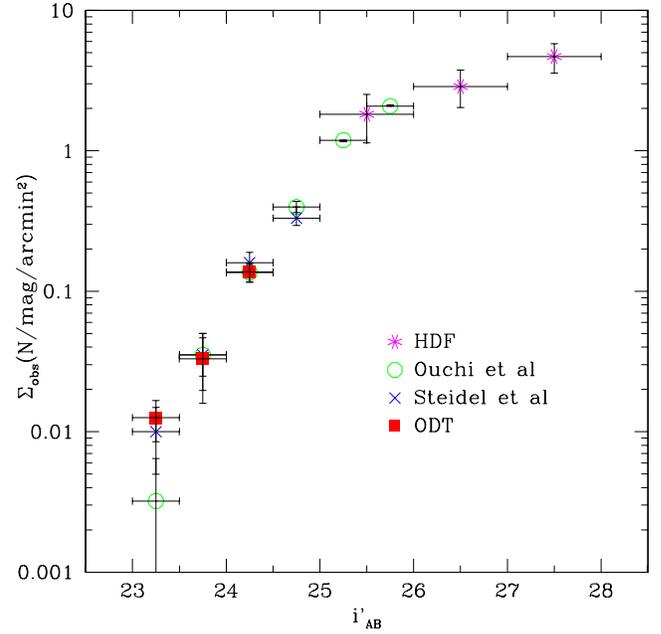


Figure 9. The surface density of colour–colour selected LBGs in the ODT survey (filled squares). Also shown are data from Ouchi et al. (2001, open circles), Steidel et al. (1999, crosses) and data from the *HDF* Arnouts et al. (1999, asterisks). Surface densities are plotted per unit magnitude.

We also calculate the space density using

$$n_{\Sigma} = A_{\Omega} \Sigma \left[\int_0^{\infty} N(z) \frac{dV}{dz} dz \right]^{-1}, \quad (7)$$

where Σ is the measured surface density in arcmin $^{-2}$, as shown in Fig. 9. We use $N(z)$ from equation (6), and $dV(z)$ is the comoving differential volume element per square steradian for the cosmology of choice. The coefficient $A_{\Omega} = 1.1818 \times 10^7$ arcmin 2 sr $^{-2}$. The calculated surface and space densities are presented in Table 6. In the next section, we also explore how sensitive the correlation function parameters are relative to the choice of $N(z)$, and to the limiting magnitude or chosen magnitude range.

5 CORRELATION FUNCTIONS

5.1 Angular correlation function

One of the more important results to come from LBG studies has been the measure of their angular correlation function, and associated r_0 and bias parameter, ‘ b ’. It has been shown (e.g. Wechsler et al. 2001; Bullock et al. 2002) that this can be used to constrain the halo occupation function for LBGs (see Section 6), which in turn provides information on their typical masses and likely evolutionary fate (e.g. Moustakas & Somerville 2002).

The clustering of galaxies, as represented by the two-point correlation function, has been shown to be well approximated by a power law (e.g. Peebles 1980) of the form $\xi(r) = (r/r_0)^{-\gamma}$. The angular projection of this also follows a power law of the form $w(\theta) = A_w \theta^{-\beta}$, where $\beta = \gamma - 1$. We calculate the two-point angular correlation function, $\omega(\theta)$, for i' -band selected galaxies, using the estimator of Landy & Szalay (1993)

$$\omega(\theta) = 1 + \frac{DD(\theta)}{RR(\theta)} W_1 - 2 \frac{DR(\theta)}{RR(\theta)} W_2, \quad (8)$$

where $DD(\theta)$, $RR(\theta)$ and $DR(\theta)$ are the numbers of ‘data–data’, ‘random–random’ and ‘data–random’ pairs, respectively, and W_1 and W_2 account for the numbers of data and random points (N_{ran} and N_{data}) used to estimate the correlation function (Roche & Eales 1999). Here,

$$W_1 = \frac{N_{\text{ran}}(N_{\text{ran}} - 1)}{N_{\text{data}}(N_{\text{data}} - 1)}, \quad (9)$$

and

$$W_2 = \frac{(N_{\text{ran}} - 1)}{N_{\text{data}}}. \quad (10)$$

In the data we remove bright stars, diffraction spikes, bad columns, etc., and constrain the random sample so that it follows the same geometrical constraints as the data.

The Poisson error for the angular correlation function can be calculated using

$$\sigma_{\omega(\theta)} = \frac{1 + \omega(\theta)}{\sqrt{DD(\theta)}}. \quad (11)$$

However, $\sigma_{\omega(\theta)}$ is probably an underestimate of the actual error in $\omega(\theta)$ when the number of data points used is small (Baugh et al. 1996). In order to obtain a more appropriate estimate of the errors on $w(\theta)$, we compute bootstrap errors (Ling, Barrow & Frenk 1986).

If a sample contains n galaxies, then a ‘bootstrap’ sample can be created by drawing (with replacement) n galaxies from the original galaxy sample. This process can be repeated N times and the correlation function, $w_i(\theta)$, calculated for each of the N bootstrap samples (where $i = 1, 2, 3, \dots, N$). The estimate of the error in $w(\theta)$ is then given by

$$\sigma_{\text{boot}} = \sqrt{\left\{ \sum_{i=1}^N \frac{[w_i(\theta) - w_{\text{av}}(\theta)]^2}{N-1} \right\}}, \quad (12)$$

The computed bootstrap errors are shown in Figs 10 and 11, and are used in what follows.

Due to the fact that we are using a finite solid angle, we correct the estimate of $\omega(\theta)$ for the integral constraint (Groth & Peebles

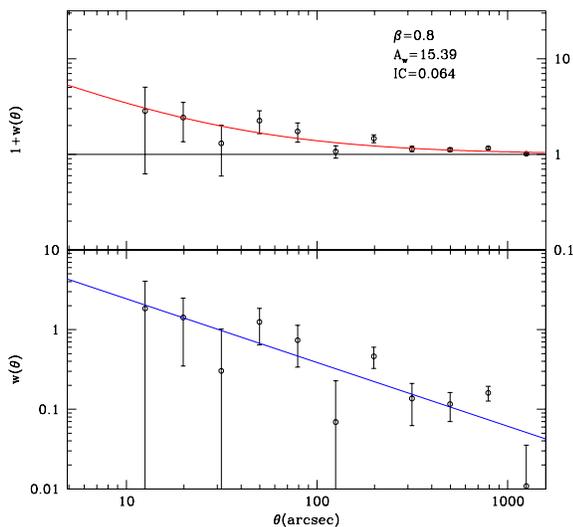


Figure 10. The angular correlation function for 66 $i' < 24.5$ LBGs (bottom panel). The fit is restricted to a fixed slope of $\beta = 0.8$. The top panel shows $1 + w(\theta)$. Data points are corrected for the integral constraint.

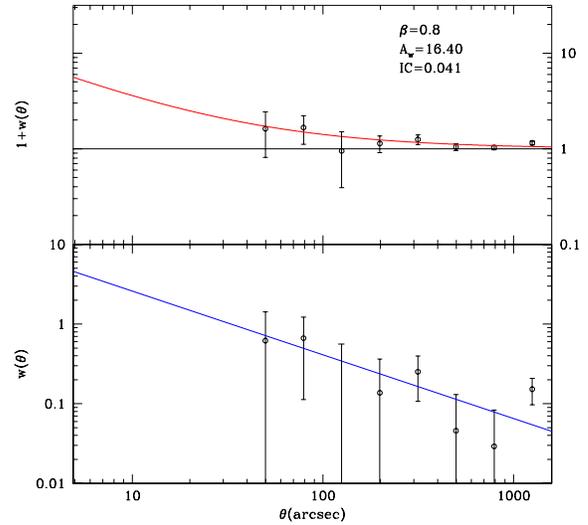


Figure 11. The angular correlation function for 74 $i' < 24.0$ LBGs (bottom panel). The fit is restricted to a fixed slope of $\beta = 0.8$. The top panel shows $1 + w(\theta)$. Data points are corrected for the integral constraint.

1977; Roche & Eales 1999)

$$IC = \frac{1}{\Omega^2} \iint \omega(\theta_{12}) \delta\Omega_1 \delta\Omega_2, \quad (13)$$

where Ω is the solid angle. The integral constraint can be estimated (providing enough random points are used) with $IC = A_w B$ where B is given by

$$B = \frac{\sum RR(\theta)\theta^{-\beta}}{\sum RR(\theta)}. \quad (14)$$

The angular correlation function then becomes

$$w(\theta) = w_{\text{obs}}(\theta) + IC = A_w \theta^{-\beta}. \quad (15)$$

The best-fitting parameters A_w and β can be found by finding the best (χ^2 minimization²) fit to the function $w_{\text{obs}}(\theta) = A_w(\theta^{-\beta} - B)$, for different values of β , where B is recalculated each time. The error on A_w is computed as the value that gives $\Delta\chi^2 = 1$.

Fig. 10 shows the angular correlation function for the fainter sample of 66 LBG candidates covering an area where we are confident there is comparable completeness to $i' = 24.5$. Some 90 000 random objects were used in this calculation. The angular correlation function for the ‘bright’ sample to $i' = 24.0$ is shown in Fig. 11 based on 74 LBG candidates. Here, 250 000 random objects were used to calculate the correlation function. Because the clustering strength of faint galaxies is significantly less than that discussed below for bright LBGs, faint galaxies in the same catalogue could also be used as an effective sample of ‘random’ objects. This provides a good test that there is no remnant structure due to varying completeness in the i' -band data. The galaxies to be used as ‘random’ objects were selected to cover the same magnitude range as the LBGs (i.e. $23.0 < i' < 24.5$ for the faint sample and $23.0 < i' < 24.0$ for the bright sample). Similar correlation strengths to using a truly random sample were obtained. For the $i' < 24.5$ sample, a slightly higher value

² The technique of χ^2 minimization assumes independent errors. In fact, the errors on $w(\theta)$ measurements are correlated. This complication is neglected here, but we note that if χ^2 is used to test goodness of fit, it is likely to be an underestimate.

Table 7. Summary of correlation function parameters, bias and number densities. The bias is calculated as $b = \sigma_g/\sigma_{\text{dm}}$ using linear theory and a power spectrum via $n = 1$ and $\sigma_8 = 0.9$.

Reference	Sample	A_w	r_0 (h_{100}^{-1} Mpc)	b	n (h_{100}^3 Mpc $^{-3}$)
ODT (this paper)	$i' < 24.0$	16.40	$11.8^{+3.1}_{-4.0}$	$8.4^{+2.0}_{-2.6}$	$(2.23 \pm 0.25) \times 10^{-5}$
ODT (this paper)	$i' < 24.5$	15.39	$11.4^{+1.7}_{-1.9}$	$8.1^{+1.1}_{-1.2}$	$(8.86 \pm 0.49) \times 10^{-5}$
Ouchi et al. (2001)	$i' < 25.5$	0.97	$3.2^{+1.0}_{-1.2}$	$2.6^{+0.7}_{-0.9}$	$(1.71 \pm 0.07) \times 10^{-3}$
Ouchi et al. (2001)	$i' < 26.0$	0.71	$2.7^{+0.5}_{-0.6}$	$2.2^{+0.4}_{-0.5}$	$(3.72 \pm 0.11) \times 10^{-3}$
Ouchi et al. (2004)	$i' < 24.8$	6.1	$7.9^{+2.1}_{-2.7}$	$5.3^{+1.3}_{-1.7}$	$(2.2 \pm 0.6) \times 10^{-4}$
Ouchi et al. (2004)	$i' < 25.3$	2.6	$5.1^{+1.1}_{-1.0}$	$3.5^{+0.6}_{-0.7}$	$(9.2 \pm 1.10) \times 10^{-4}$
Ouchi et al. (2004)	$i' < 26.0$	1.7	$4.1^{+0.2}_{-0.2}$	$2.9^{+0.1}_{-0.1}$	$(4.9 \pm 0.30) \times 10^{-3}$

of $A_w = 15.52$ was measured ($\chi_{\text{red}}^2 = 1.01$). For $i' < 24.0$, a value of $A_w = 17.58$ was obtained ($\chi_{\text{red}}^2 = 0.89$).

There is increasing evidence that faint samples of I -band selected galaxies (and therefore galaxies with a higher median redshift) have a shallower slope, β , to the best-fitting power law $w(\theta) = A_w \theta^{-\beta}$ than the often quoted $\beta = 0.8$ seen in clustering studies of local galaxies (e.g. Brainerd & Smail 1998; Postman et al. 1998; McCracken et al. 2001). In addition, many measurements of LBG clustering also find a best-fitting slope that is shallower than 0.8 (e.g. Ouchi et al. 2001; Porciani & Giavalisco 2002), although few of the measurements are accurate enough to test whether or not this is significantly different to the local Universe, and so results are usually quoted with a fit constrained to $\beta = 0.8$.

In the ODT data for $i' < 24.5$, the best-fitting parameters (with θ in arcsec) are consistent with this ($A_w = 7.43^{+12.4}_{-5.0}$, $\beta = 0.63^{+0.24}_{-0.23}$). However, there are not enough data to place a strong constraint on this slope. Therefore, to be consistent with other measures of LBG clustering in the literature, only results for a constrained slope of $\beta = 0.8$ are considered in what follows. However, we note that LBG correlation functions with a shallower slope would have the general effect of increasing estimates of r_0 (in this case by ~ 15 per cent). This is something that will need to be looked into in more detail in the full ODT sample, and in other high-quality data sets probing the high-redshift Universe.

Confining the slope of the angular correlation function to 0.8, we obtain an amplitude A_w of 15.39 ± 4.3 arcsec $^{0.8}$ for the faint $i' < 24.5$ sample. For the brighter sample, a best-fitting amplitude of $A_w = 16.40 \pm 8.6$ arcsec $^{0.8}$ was calculated. The quoted errors in A_w correspond to the error on the best fit to the power law $w(\theta) = A_w \theta^{-0.8}$. We make no attempt here to correct the correlation function for contamination, in line with similar studies with similar contamination estimates. However, we note that, in the case of maximum contamination, our estimates of A_w could be too low by ~ 20 per cent.³ The effect on the deprojected r_0 , as we discuss in the next section, will be smaller.

5.2 Spatial correlation function

Using our adopted redshift distribution, $N(z)$ (Section 4.2), in conjunction with the best-fitting parameters to our angular correlation

³ If the contaminants are unclustered then A_w would be reduced by a factor $(1 - f)^2$, where f is the contamination fraction. In the case where contaminants are clustered, and their clustering amplitude is smaller than that for LBGs, the reduction in A_w would be less.

functions, we estimate the corresponding correlation lengths for galaxies using the Limber deprojection (Peebles 1980; Efstathiou et al. 1991). The angular correlation function, $w(\theta)$, and spatial correlation function, $\xi(r)$, are related via

$$A_w = r_0^\gamma C \frac{\int_0^\infty F(z) D_\theta^{1-\gamma}(z) N(z)^2 g(z) dz}{\left[\int_0^\infty N(z) dz \right]^2}, \quad (16)$$

where C is a constant

$$C = \sqrt{\pi} \frac{\Gamma[(\gamma - 1)/2]}{\Gamma(\gamma/2)}, \quad (17)$$

$D_\theta(z)$ is the angular diameter distance, $g(z)$ is a function that describes the cosmology

$$g(z) = \frac{H_0}{c} \left\{ (1+z)^2 [1 + \Omega_m z + \Omega_\Lambda (1+z)^{-2} - 1]^{1/2} \right\}, \quad (18)$$

and $F(z)$ is the redshift dependence of $\xi(r)$.

Because the selection function for LBGs is very narrow (especially as a function of time), the function $F(z)$ can be removed from the integral. This can be verified by using both linear and non-linear clustering evolution for the function $F(z)$ over our redshift distribution. We find this gives only negligible changes in r_0 (cf. Ouchi et al. 2001; Porciani & Giavalisco 2002). The comoving r_0 at redshift z is therefore given by

$$r_0(z) = r_0 [F(z)]^{1/\gamma}. \quad (19)$$

Using our fiducial redshift distribution and the angular correlation function parameters derived in Section 5.1 we obtain a spatial correlation length (at $z \sim 4$) of $r_0 = 11.4^{+1.7}_{-1.9} h_{100}^{-1}$ Mpc for the $i' < 24.5$ sample. For the $i' < 24.0$ sample, we obtain $r_0 = 11.8^{+3.1}_{-4.0} h_{100}^{-1}$ Mpc (see Table 7 for a summary of these results). These correlation lengths are considerably greater than other measurements at $z \sim 4$ (e.g. Arnouts et al. 1999, 2002; Ouchi et al. 2001) and the implications and context of this are discussed in Section 6.

In order to test the accuracy of these results we vary the $N(z)$ distribution introduced in Section 4.2 over a reasonable range of values for \bar{z} and σ_z and also consider a simple top-hat redshift distribution (see Fig. 8). We find that the calculated r_0 is only weakly dependent on the \bar{z} used, and is only affected at the 10 per cent level, even using a top-hat distribution over $3.5 < z < 4.5$. However, significant broadening of the selection function by increasing σ_z to 0.4 would increase r_0 by up to 40 per cent. It should also be noted that a narrower selection function would lead to smaller values of r_0 . A Gaussian selection function with σ_z reduced to 0.1 would lead to an r_0 that is ~ 20 per cent smaller. If we consider the maximum final contamination given in Table 3, and assume that contaminants are unclustered (see Section 5.1), we find that r_0 would increase by

15 per cent. Note that our quoted results contain the error derived from the best-fitting power law only. In line with other LBG studies, the effects of systematic uncertainties are not considered further.

5.3 Linear bias

The spatial correlation length r_0 can be used to calculate the linear bias. Several definitions of linear bias appear in the literature; here we adopt the definition based on the variation in mass over $8 h^{-1}$ Mpc spheres (σ_8) and bias is defined at our fiducial median redshift ($\bar{z} = 4.0$).

Using our correlation function measurements, and assuming a Λ CDM cosmology with an $n = 1$ power spectrum of initial mass fluctuations, normalized to $\sigma_8 = 0.9$, we use linear theory (neglecting any non-linear corrections which are expected to be small at $z \sim 4$) to determine the linear bias

$$b = \frac{\sigma_g}{\sigma_{\text{dm}}}, \quad (20)$$

all measured at a scale of $r = 8 h_{100}^{-1}$ Mpc.

In linear theory, the dark matter variance, σ_{dm} , at redshift $z = 0$ can be calculated for an arbitrary mass scale M , which is equivalent to a physical scale R , via

$$R = \left(\frac{3}{4\pi} \frac{M}{\rho_b} \right)^{1/3}, \quad (21)$$

where the density, $\rho_b = \Omega_m (2.78 \times 10^{11}) h_{100}^2$. The dark matter variance, $\sigma_{\text{dm}}(z = 0)$, can then be calculated for a given physical scale R , using the relation between mass scale and variance. At redshift z , the dark matter variance is given by

$$\sigma_{\text{dm}}(z) = \sigma_{\text{dm}}(z = 0) D_{\text{lin}}(z), \quad (22)$$

where $D_{\text{lin}}(z)$ is the linear growth factor at the redshift of interest; at redshift 4, $D_{\text{lin}}(z = 4) = 0.25569$. σ_g is calculated (assuming ξ is described by a power law) as

$$\sigma_g^2 = J_2 \left(\frac{r_0}{r} \right)^\gamma, \quad (23)$$

where

$$J_2 = \frac{72}{(3 - \gamma)(4 - \gamma)(6 - \gamma)2^\gamma}, \quad (24)$$

(Peebles 1980). See Somerville et al. (2004) for further details and discussion.

For $i' < 24.5$, we obtain $b = 8.1_{-1.2}^{+1.1}$, and for $i' < 24.0$, $b = 8.4_{-2.6}^{+2.0}$. The clustering and bias results are summarized in Table 7.

6 DISCUSSION

In hierarchical models of structure formation, structure forms by the magnification of initial density fluctuations by gravitational instability. A key feature is that virialized structures (or dark matter haloes) should have clustering properties that differ from that of the overall mass distribution, with more massive haloes being more strongly clustered (e.g. Kaiser 1984). If galaxies and clusters form when baryonic material falls into the potential wells of the dark matter haloes, then a correlation between galaxy (and hosting halo) mass and clustering strength (and therefore bias) would be expected.

An important measurement in studies of LBGs has been that of their clustering properties. The strong spatial clustering and surface densities exhibited by LBGs at $z \sim 3$ appears to be consistent with biased galaxy formation, implying an association between LBGs and

fairly massive dark matter haloes, even if there are several galaxies per dark matter halo (e.g. Giavalisco et al. 1998; Giavalisco & Dickinson 2001; Somerville et al. 2001; Wechsler et al. 2001; Bullock et al. 2002). The measurement of the clustering properties of $z \sim 4$ LBGs from the ODT survey provides values of r_0 and b that are significantly larger than previous measurements of these parameters at this redshift (Ouchi et al. 2001; Arnouts et al. 1999), and of U -band dropouts at $z \sim 3$ (Giavalisco et al. 1998; Arnouts et al. 1999; Giavalisco & Dickinson 2001; Porciani & Giavalisco 2002). However, the data presented here are from a sample that is much brighter than other measurements of the correlation function at $z \sim 4$. We are therefore able, by combining these shallower wide-field data with deeper pencil-beam surveys such as the *HDF* and the Subaru Deep Field, to compare with the luminosity-dependent bias detected in the local Universe (Norberg et al. 2002) and at high redshifts (Giavalisco & Dickinson 2001; Foucaud et al. 2003; Ouchi et al. 2004), and test models of biased galaxy formation. Using our definition of bias (equation 20) we calculate bias parameters for the $z \sim 4$ Subaru deep field clustering results of Ouchi et al. (2001). We obtain $b = 2.6$ for $i' < 25.5$ and $b = 2.2$ for $i' < 26.0$. We also consider the more recent results of Ouchi et al. (2004), who present clustering results for several differently defined LBG samples. The three samples using data up to a given magnitude limit (with $b = 5.3$ for $i' < 24.8$, $b = 3.5$ for $i' < 25.3$ and $b = 2.9$ for $i' < 26.0$) are used here. When these data are considered alongside the data presented in this paper, there appears to be a clear trend between the depth of a sample and LBG bias. At $z \sim 4$, our faintest magnitude limit of $i' \approx 24.5$ corresponds to a rest-frame absolute magnitude of $M_{1700} \approx -20.5$ at $\lambda_{\text{rest}} = 1700 \text{ \AA}$.⁴ Employing the $z \sim 4$ luminosity function of Steidel et al. (1999), the faintest galaxies in our sample are then $L_{1700} \approx 1.5L_*$ (the brightest being $L_{1700} \approx 6.1L_*$), whereas the Ouchi et al. (2001) and Arnouts et al. (1999, 2002) samples are, on the whole, $\text{sub}L_*$. The luminosity/bias results are summarized in Table 7 and are plotted in Fig. 12.

One interpretation of luminosity-dependent bias could be a tight correlation between dark matter halo mass and star formation rate (i.e. ultraviolet luminosity; Giavalisco & Dickinson 2001; Steidel et al. 1998). The observed i' -band magnitudes measured here correspond to a rest frame of $\lambda \sim 1500 \text{ \AA}$. At this wavelength we are measuring the ultraviolet luminosity, which is a good tracer of the star formation rate in the galaxy. We can therefore infer a link between halo mass and star formation rate. However, we emphasize that the rest-frame ultraviolet luminosity is more indicative of the instantaneous star formation rate, rather than total underlying light (and therefore stellar mass).

6.1 Masses and environments of bright $z \sim 4$ LBGs

In Fig. 13 we plot bias versus comoving space density for galaxies at $z \sim 4$. The observed space densities are calculated in Section 4.3. For the ODT bright sample of $i' < 24.0$ and the faint sample of $i' < 24.5$, we obtain bias values of 8.4 and 8.1, respectively. The corresponding space densities, calculated using the same selection function used for the correlation function inversion, are 2.23×10^{-5} and $8.86 \times 10^{-5} h_{100}^3 \text{ Mpc}^{-3}$. We also show points for the Ouchi et al. (2004) and Ouchi et al. (2001) results, using their estimated effective volume and the reported numbers of LBGs for each limiting magnitude (with estimated uncertainties).

⁴ To calculate the appropriate k -corrections, we have used the rest-frame ultraviolet SED of Shapley et al. (2003).

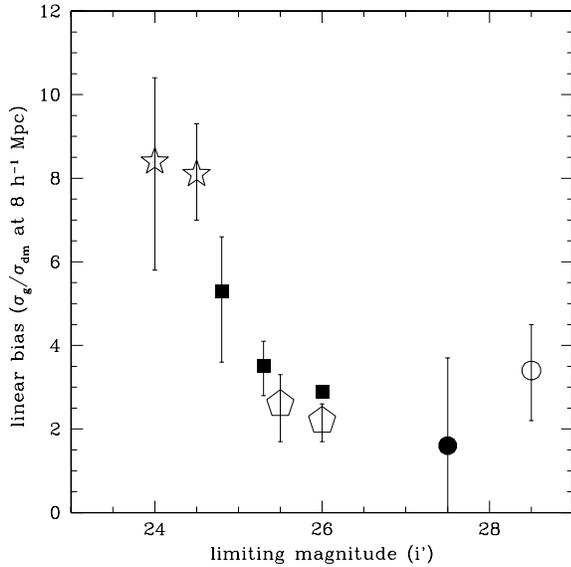


Figure 12. The relationship between the limiting i' -band magnitude of a survey and the measured bias (using the σ_8 definition as discussed in the text). ODT data are represented by large open stars. Data from Ouchi et al. (2001) are shown as open pentagons. Filled squares correspond to data from Ouchi et al. (2004). The filled circle represents results from Arnouts et al. (1999), and the open circle shows results from Arnouts et al. (2002).

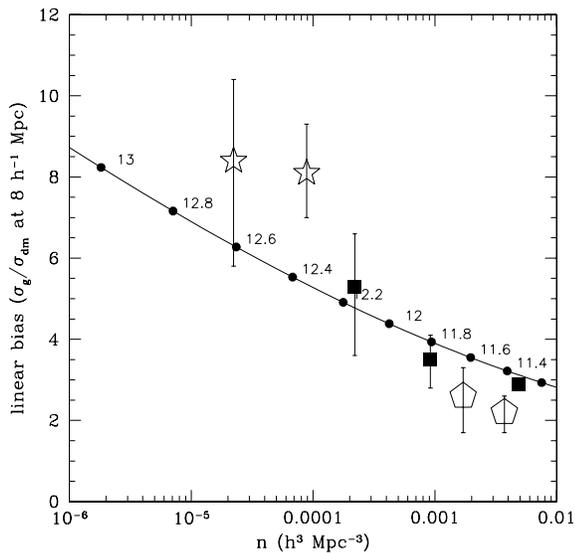


Figure 13. The relationship between galaxy number density and the linear bias parameter, for a simple galaxy halo occupation function with $\alpha = 0$ (i.e. one galaxy per halo) is shown by the solid black line. The numbers marked on the line correspond to the dark matter halo masses hosting the observed galaxies in this model. Masses are written as $\log_{10}(M)$ in units of $h^{-1} M_{\odot}$. Open stars represent ODT data presented in this paper. Open pentagons represent the results of Ouchi et al. (2001). Filled squares correspond to data from Ouchi et al. (2004). See Table 7 for a summary of these values.

The line in Fig. 13 depicts the correspondence between linear bias and comoving space density for dark matter haloes at $z = 4$ (based on Sheth & Tormen 1999; see also Somerville et al. 2004). Smaller space densities correspond to more rare overdensities, and therefore to higher minimum dark matter halo masses (as shown). This is the relation that Λ CDM predicts that galaxies would follow, if there

were only one galaxy in each dark matter halo, which is a useful simplification for checking trends, but not how galaxies tend to exist. In cases where there may be varying numbers of galaxies per halo, particularly likely if these are starbursting galaxies, the ‘halo occupation’ formalism is a useful parametrization. Here, above some threshold minimum mass, we assume there is a function $N(M > M_{\min}|\alpha, M_1)$ that describes both the internal mass-function slope (α), and the mass of a dark matter halo that typically hosts one galaxy (M_1); therefore, there may statistically be $M > M_{\min}$ haloes that have no resident galaxies). This prescription is found to describe galaxies at both $z \sim 0$ very well (Marinoni & Hudson 2002; van den Bosch, Yang & Mo 2003; Magliocchetti & Porciani 2003) and $z \sim 3$ (e.g. Wechsler et al. 2001; Bullock et al. 2002; Kravtsov et al. 2004).

A detailed exploration of the occupation statistics of LBGs implied by our measurements is beyond the scope of this paper. It is useful, however, to turn to the one-galaxy-per-halo curve, and note two items. First, the data span a large dynamic range of space densities, and yet follow the same general trend as the haloes. This suggests that the generic b/n correlation predicted by Λ CDM is reflected in the data. Secondly, the mass scale of the ODT LBG sample, $M_{\min} \sim \text{few} \times 10^{12} h^{-1} M_{\odot}$, begs the question whether these naturally connect with later galaxy populations known to be in similarly massive environments. In Moustakas & Somerville (2002), the possible clustering evolution between massive galaxy populations at $z \sim 0, 1.2$, and 3 were explored in the context of simple Λ CDM-motivated models. The $z \sim 3$ galaxy population characteristics were drawn from the literature of largely sub L_* samples, and were shown to not have a natural connection with the lower-redshift ($z \sim 1.2$) populations, which are plausibly connected with the most massive (elliptical galaxy type) environments today. The $z \sim 1.2$ (extremely red object) population’s linear bias and comoving space density, when extrapolated to $z \sim 4$ (under the ‘constant minimum mass’ model; Moustakas & Somerville 2002) are expected to have $b \approx 8$ and $n \approx 10^{-5} h^3 \text{Mpc}^{-3}$ – values remarkably close to those measured in the super L_* ODT LBG sample.

6.2 Significance of results

Thus far, the quoted errors on our measured r_0 and bias have only included ‘random’ errors based on the best fit to the angular correlation function. In Section 4.2, we derived our redshift distribution based on an analysis of theoretical colours of high-redshift galaxies. We also considered a range of possible redshift distributions and found that this led to a 10 per cent uncertainty on the derived value for r_0 . Combining this error in quadrature with the measured statistical errors leads to $r_0 = 11.4^{+2.0}_{-2.2}$ and $r_0 = 11.8^{+3.3}_{-4.1}$. This is a difference from the results of Ouchi et al. (2001) at the 3.7σ level for the faint sample and 2.1σ for the bright sample.

We also noted in Section 5.1 that the effect on contamination on our LBG sample is likely to dilute the correlation function, and any correction that might be applied would lead to an increase in r_0 , further strengthening any relationship between luminosity and bias.

6.2.1 Cosmic variance

Because the LBG population is known to be strongly clustered, it is clear that the number density of objects will vary from field to field. It is therefore important that a large enough volume of sky is surveyed so that this ‘cosmic variance’ does not significantly bias the measured number density of objects relative to the overall cosmic average. In order to estimate the size of such an effect on this data

set, we use the results of Somerville et al. (2004) who compute the variance of dark matter from linear theory as a function of redshift.

Based on the estimated effective volume at $z \sim 4$ covered in this survey ($3 \times 10^4 \text{ Mpc}^3$, assuming $h = 0.7$), the variance in the mean dark matter density is estimated to be 15 per cent (see fig. 3b in Somerville et al. 2004). How cosmic variance relates to an uncertainty in number density of a population depends on how biased that population is relative to the overall dark matter distribution. Assuming the bias measured in Section 5.3 is correct, then the uncertainty in the number density of LBGs due to cosmic variance can be estimated as $\sigma_{\text{cv}} = 8.1 \times 0.15 = 1.2$. Therefore, in Fig. 13 there is an additional uncertainty in the estimated space density due to the sample size (volume) and cosmic variance associated with this.

Could fluctuations in number density caused by cosmic variance produce the observed difference in clustering strength between the ODT and fainter samples? This is difficult to test because estimates of the size of cosmic variance effects from linear theory require prior knowledge of the relationship between the galaxy population being studied and the underlying dark matter distribution (i.e. bias). However, if we start with the null hypothesis that the intrinsic bias of the ODT LBG sample is actually the same as that measured by Ouchi et al. (2001), then the uncertainty in number density due to cosmic variance, $\sigma_{\text{cv}} = 2.6 \times 0.15 = 0.39$ (following Somerville et al. 2004). If we have underestimated the intrinsic number density due to cosmic variance, and assume that $\xi \propto 1/(n^2)$, then the ‘true’ number density could indeed be larger, and the correlation length (and hence bias) could be smaller.⁵ Taking $\sigma_{\text{cv}} = 0.39$, and assuming a 2σ fluctuation, implies an upper limit on the number density of $1.58 \times 10^{-4} h_{100}^3 \text{ Mpc}^{-3}$. If this corresponds to the Ouchi et al. (2001) correlation length of $r_0 = 3.2 h_{100}^{-1} \text{ Mpc}$, then our measured value for the number density ($8.86 \times 10^{-5} h_{100}^3 \text{ Mpc}^{-3}$) would scale r_0 to $6.1 h_{100}^{-1} \text{ Mpc}$ which is inconsistent with our measured values (in fact, a 5.5σ fluctuation in number density is required to scale the Ouchi et al. (2001) value to $11.4 h_{100}^{-1} \text{ Mpc}$). Variations in number density due to cosmic variance are too small to explain the difference in clustering strength between the ODT and fainter samples. However, it is clear that wide-field surveys at least as large as the ODT are required to measure the clustering properties of the brightest objects. We also note that the independent results of Ouchi et al. (2004) are consistent with the trend of increasing clustering strength with brighter magnitudes at $z \sim 4$.

6.3 Small-scale clustering and close pairs

In addition to a luminosity-dependent bias, Porciani & Giavalisco (2002) provide evidence from a fairly bright sample of $z \sim 3$ LBGs that there is also a significant scale-dependent bias. In their data set, there appears to be a lack of power in the angular correlation function (and therefore a lower bias) on scales less than 30 arcsec. Bullock et al. (2002) also demonstrate (where redshifts are available) how close pair statistics can be used within the halo occupation function formalism to place constraints on different models for the occupation function. Porciani & Giavalisco (2002) interpret the ‘break’ in their

⁵ This approximation, $\xi \propto 1/(n^2)$, assumes that Gaussian statistics apply, which is not the case here. A full calculation of the effects of cosmic variance requires knowledge of higher-order clustering statistics. Simulations by Foucaud et al. (2003) suggest that for a similar bright sample but at $z \sim 3$, the size of the cosmic error is likely to be smaller than the measured Poisson errors. However, this rough calculation provides a useful measure of the potential size of the ‘cosmic error’.

correlation function as evidence that the dark matter haloes have a fairly large size and mass, which is more consistent with one-galaxy-per-halo occupation functions. However, it should be noted that our analysis (and that of most other LBG clustering studies) has assumed a simple power law for the correlation function with $\gamma = 1.8$. In addition, the results of Kravtsov et al. (2004) (which come from N -body simulations) indicate that the correlation function may actually steepen on the smallest scales (which, in turn, would lead to overestimates of r_0 for LBGs). The models of Hamana et al. (2004) indicate that the observed angular correlation function is well fitted by a two-component power law with a steep component dominating on small scales, although Ouchi et al. (2004) fit the same data with single-component model with a relatively steep ($\beta = 0.9$) slope.

In the ODT data presented here, we find no close pairs on scales less than 10 arcsec in the $i' < 24.5$ sample and none less than 40 arcsec in the brighter $i' < 24.0$ sample. Unfortunately, because only one or less pairs would be required to fit the best-fitting power law on these scales, it is not possible to determine whether or not this constitutes evidence to support the results of Porciani & Giavalisco (2002), or indeed to determine whether a steeper slope is more appropriate on small scales. However, an analysis of the full ODT survey should yield a good measurement of the small-scale clustering of bright LBGs, and help assess the significance of any lack of power on small scales in the angular correlation function.

7 CONCLUSIONS

The main results presented here can be summarized as follows.

(i) The clustering of LBGs at $z \sim 4$ is well approximated by a power law $w(\theta) = A_w \theta^{-\beta}$. Fixing the slope to that of local galaxies ($\beta = 0.8$) provides a good fit to the data with $A_w = 15.39 \text{ arcsec}^{0.8}$ the best amplitude for $i' < 24.5$ and $A_w = 16.40 \text{ arcsec}^{0.8}$ for $i' < 24.0$. However, we note that there is evidence to suggest that $\beta = 0.8$ may not be a valid assumption to make at high redshift, and a shallower slope may be more appropriate.

(ii) Using a reasonable fiducial redshift distribution to characterize the selection function for LBGs, the angular correlation function can be deprojected to obtain the spatial correlation length, r_0 . For our fainter sample ($i' < 24.5$) we obtain $r_0 = 11.4_{-1.9}^{+1.7} h_{100}^{-1} \text{ Mpc}$. Using a slightly brighter sample ($i' < 24.0$) we obtain a similar correlation length of $r_0 = 11.8_{-4.0}^{+3.1} h_{100}^{-1} \text{ Mpc}$. Comparing these results to the clustering properties of dark matter, we obtain linear bias values of $8.1_{-1.2}^{+1.1}$ and $8.4_{-2.6}^{+2.0}$, respectively.

(iii) When compared with fainter surveys, the bias and correlation lengths seen in the ODT are clearly significantly larger. We interpret this as evidence for a luminosity-dependent bias for LBGs, as predicted by some semi-analytical models. The ODT bias values are shown to be consistent with a simple model using a galaxy occupation function describing one observable galaxy per dark matter halo. Bright (super L_*) LBGs seem to be more biased tracers of mass than fainter (sub L_*) ones, suggesting a relationship between halo mass and instantaneous star formation rate.

(iv) The population of sub L_* LBGs is unlikely to be the progenitor of massive galaxies at $z \sim 1$ and $z \sim 0$. However, the extremely bright, super L_* population presented in this paper has biases, number densities and halo masses ($M_{\text{min}} \sim \text{few} \times 10^{12} h^{-1} M_{\odot}$) that are consistent, in a simple model, with them being potential progenitors of the most massive galaxies.

The data presented here are only a small sample of the full ODT survey. With an extended sample containing >1000 LBGs over a much wider area, it should be possible to place stronger constraints

on the clustering properties of this bright sample of high-redshift galaxies, and address issues such as their scale-dependent bias. In addition, other surveys currently underway, such as the National Optical Astronomy Observatory (NOAO) Deep Wide-Field Survey (NDWFS; Jannuzi & Dey 1999), the Great Observatories Origins Deep Survey (GOODS; Dickinson et al. 2003), the VIRMOS-VLT Deep Survey (Le Fevre et al. 1998), the Canada–France–Hawaii Telescope (CFHT) Legacy Survey,⁶ Subaru/*XMM-Newton* Deep Survey (SXDS; Kodama et al. 2004) and COSMOS,⁷ should provide good measurements of LBG clustering probing (when combined) a large dynamic range in luminosity and redshift. This should allow tighter constraints on parameters such as the galaxy occupation function, and the scale- and luminosity-dependent bias for LBGs; thus, providing a better understanding of the relationship between LBGs, other high-*z* populations, and galaxies today.

ACKNOWLEDGMENTS

The INT is operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. PDA, CEH and CAB acknowledge the support of UK Particle Physics and Astronomy Research Council (PPARC) studentships. ECM thanks the C. K. Marr Educational Trust. This work was supported by the PPARC Rolling Grant PPA/G/O/2001/00017 at the University of Oxford. LAM acknowledges support from the *SIRTF* Legacy Science Program, provided through an award issued by the Jet Propulsion Laboratory, California Institute of Technology under NASA contract 1407. The authors thank both anonymous referees whose comments helped to improve this paper.

REFERENCES

Adelberger K. L., Steidel C. C., 2000, *ApJ*, 544, 218
 Adelberger K. L., Steidel C. C., Giavalisco M., Dickinson M., Pettini M., Kellogg M., 1998, *ApJ*, 505, 18
 Adelberger K. L., Steidel C. C., Shapley A. E., Pettini M., 2003, *ApJ*, 584, 45
 Arnouts S., Cristiani S., Moscardini L., Matarrese S., Lucchin F., Fontana A., Giallongo E., 1999, *MNRAS*, 310, 540
 Arnouts S. et al., 2002, *MNRAS*, 329, 355
 Bahcall J. N., 1986, *ARA&A*, 24, 577
 Baugh C. M., Gardner J. P., Frenk C. S., Sharples R. M., 1996, *MNRAS*, 283, L15
 Baugh C. M., Cole S., Frenk C. S., Lacey C. G., 1998, *ApJ*, 498, 504
 Bell E. F. et al., 2004, *ApJ*, 608, 752
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Brainerd T. G., Smail I., 1998, *ApJ*, 494, L137
 Bullock J. S., Wechsler R. H., Somerville R. S., 2002, *MNRAS*, 329, 246
 Coleman G. D., Wu C.-C., Weedman D. W., 1980, *ApJS*, 43, 393
 de Lapparent V., Galaz G., Bardelli S., Arnouts S., 2003, *A&A*, 404, 831
 Dickinson M., Giavalisco M., The Goods Team, 2003, in *Proc. ESO Workshop, The Mass of Galaxies at Low and High Redshift*, Venice, Italy, 24–26 October 2001. Springer-Verlag, Berlin, p. 324
 Efsthathiou G., Bernstein G., Tyson J. A., Katz N., Guhathakurta P., 1991, *ApJ*, 380, L47
 Fan X. et al., 2001, *AJ*, 121, 54
 Fioc M., Rocca-Volmerange B., 1997, *A&A*, 326, 950

Foucaud S., McCracken H. J., Le Fèvre O., Arnouts S., Brodwin M., Lilly S. J., Crampton D., Mellier Y., 2003, *A&A*, 409, 835
 Gardner J. P., 1998, *PASP*, 110, 291
 Giavalisco M., Dickinson M., 2001, *ApJ*, 550, 177
 Giavalisco M., Steidel C. C., Adelberger K. L., Dickinson M. E., Pettini M., Kellogg M., 1998, *ApJ*, 503, 543
 Glazebrook K., Peacock J. A., Collins C. A., Miller L., 1994, *MNRAS*, 266, 65
 Groth E. J., Peebles P. J. E., 1977, *ApJ*, 217, 385
 Hamana T., Ouchi M., Shimasaku K., Kayo I., Suto Y., 2004, *MNRAS*, 347, 813
 Hill G. J., Rawlings S., 2003, *NewAR*, 47, 373
 Jannuzi B. T., Dey A., 1999, in *ASP Conf. Ser. Vol. 193, The Hy-Redshift Universe: Galaxy Formation and Evolution at High Redshift*. Astron. Soc. Pac., San Francisco, p. 258
 Kaiser N., 1984, *ApJ*, 284, L9
 Kodama T. et al., 2004, *MNRAS*, 350, 1005
 Kravtsov A. V., Berlind A. A., Wechsler R. H., Klypin A. A., Gottlöber S., Allgood B., Primack J. R., 2004, *ApJ*, 609, 35
 Landolt A. U., 1992, *AJ*, 104, 340
 Landy S. D., Szalay A. S., 1993, *ApJ*, 412, 64
 Le Fevre O. et al., 1998, in *Proc. 14th IAP Meeting, Wide Field Surveys in Cosmology*, Paris, France, 1998 May 26–30. Editions Frontières, Gif-sur-Yvette, p. 327
 Ling E. N., Barrow J. D., Frenk C. S., 1986, *MNRAS*, 223, 21P
 McCracken H. J., Le Fèvre O., Brodwin M., Foucaud S., Lilly S. J., Crampton D., Mellier Y., 2001, *A&A*, 376, 756
 MacDonald E. C. et al., 2004, *MNRAS*, 352, 1255
 Madau P., 1995, *ApJ*, 441, 18
 Magliocchetti M., Porciani C., 2003, *MNRAS*, 346, 186
 Marinoni C., Hudson M. J., 2002, *ApJ*, 569, 101
 Moustakas L. A., Somerville R. S., 2002, *ApJ*, 577, 1
 Norberg P. et al., 2002, *MNRAS*, 332, 827
 Olding E. J., 2002, D.Phil. thesis, University of Oxford
 Ouchi M. et al., 2001, *ApJ*, 558, L83
 Ouchi M. et al., 2004, *ApJ*, 611, 685
 Peebles P. J. E., 1980, *The Large-Scale Structure of the Universe*. Princeton Univ. Press, Princeton, NJ
 Pettini M., Shapley A. E., Steidel C. C., Cuby J., Dickinson M., Moorwood A. F. M., Adelberger K. L., Giavalisco M., 2001, *ApJ*, 554, 981
 Pickles A. J., 1998, *PASP*, 110, 863
 Porciani C., Giavalisco M., 2002, *ApJ*, 565, 24
 Postman M., Lauer T. R., Szapudi I., Oegerle W., 1998, *ApJ*, 506, 33
 Robin A., Creze M., 1986, *A&A*, 157, 71
 Robin A. C., Haywood M., Creze M., Ojha D. K., Bienayme O., 1996, *A&A*, 305, 125
 Roche N., Eales S. A., 1999, *MNRAS*, 307, 703
 Shapley A. E., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., Pettini M., 2001, *ApJ*, 562, 95
 Shapley A. E., Steidel C. C., Pettini M., Adelberger K. L., 2003, *ApJ*, 588, 65
 Sheth R. K., Tormen G., 1999, *MNRAS*, 308, 119
 Somerville R. S., Primack J. R., Faber S. M., 2001, *MNRAS*, 320, 504
 Somerville R. S., Lee K., Ferguson H. C., Gardner J. P., Moustakas L. A., Giavalisco M., 2004, *ApJ*, 600, L171
 Steidel C. C., Hamilton D., 1992, *AJ*, 104, 941
 Steidel C. C., Giavalisco M., Pettini M., Dickinson M., Adelberger K. L., 1996, *ApJ*, 462, L17
 Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., Pettini M., Kellogg M., 1998, *ApJ*, 492, 428
 Steidel C. C., Adelberger K. L., Giavalisco M., Dickinson M., Pettini M., 1999, *ApJ*, 519, 1
 Stevens R., Lacy M., 2001, *MNRAS*, 325, 897
 van den Bosch F. C., Yang X., Mo H. J., 2003, *MNRAS*, 340, 771
 Wechsler R. H., Somerville R. S., Bullock J. S., Kolatt T. S., Primack J. R., Blumenthal G. R., Dekel A., 2001, *ApJ*, 554, 85

⁶ See <http://www.cfht.hawaii.edu/Science/CFHTLS>.

⁷ See <http://www.astro.caltech.edu/~cosmos>.

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