
7-10-2004

The Clustering of Galaxies Around Three $z \sim 3$ Damped Ly α Absorbers

Nicolas Bouché
University of Massachusetts Amherst

James D. Lowenthal
Smith College, jlowenth@smith.edu

Follow this and additional works at: https://scholarworks.smith.edu/ast_facpubs



Part of the [Astrophysics and Astronomy Commons](#)

Recommended Citation

Bouché, Nicolas and Lowenthal, James D., "The Clustering of Galaxies Around Three $z \sim 3$ Damped Ly α Absorbers" (2004). Astronomy: Faculty Publications, Smith College, Northampton, MA.
https://scholarworks.smith.edu/ast_facpubs/42

This Article has been accepted for inclusion in Astronomy: Faculty Publications by an authorized administrator of Smith ScholarWorks. For more information, please contact scholarworks@smith.edu

THE CLUSTERING OF GALAXIES AROUND THREE $Z \sim 3$ DAMPED LY-ALPHA ABSORBERSNICOLAS BOUCHÉ¹

Dept. of Astronomy, University of Massachusetts-Amherst, Amherst, MA 01003 USA
 Max Planck Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany
 European Southern Observatory, Karl-Schwarzschild-Str 2, D-85748 Garching; nbouche@eso.org

JAMES D. LOWENTHAL¹

Five College Astronomy Dept., Smith College, Northampton, MA 01063 USA;
 james@earth.ast.smith.edu

Accepted for publication in ApJ, issue 10 July 2004

ABSTRACT

We present our results on the cross-correlation of Lyman break galaxies (LBGs) around three damped Ly α absorbers (DLAs) at $z_{\text{abs}} \simeq 3$ from deep ($\mu_{I,AB}(\text{sky}) \simeq 27.6\text{mag arcsec}^{-2}$) *UBVI* KPNO 4m/MOSAIC images. The large area of the MOSAIC images, 0.31 deg^2 or $\sim 65 \times 65 h_{71}^{-1}$ Mpc co-moving at redshift $z = 3$, allows us to probe the clustering of LBGs on scales up to 20 Mpc co-moving. Our survey covers a total of 1 deg^2 and contains $\sim 3,000$ LBGs with photometric redshifts between 2.8 and 3.5. Using the redshift likelihood distributions with m_I as a prior, we selected LBGs within a redshift slice of width $W_z = 0.15$ (corresponding to σ_z , the uncertainty in photometric redshifts) centered on the redshift of the absorbers. Within that redshift slice, we find that the DLA-LBG cross-correlation w_{dg} is $w_{dg} = (1.62 \pm 1.32) \times w_{gg}$, where w_{gg} is the LBG auto-correlation. This corresponds to a correlation length of $r_o = 5 \pm 4.5 h^{-1}$ (co-moving) (or $r_o = 7 \pm 6.8 h_{71}^{-1}$ Mpc). The cross-correlation is most significant on scales 5 – 10 Mpc. Through Monte Carlo simulations, we find that w_{dg} is significantly greater than zero at the $> 95\%$ level. In three other redshift slices that do not contain a DLA, we do not find any evidence of clustering. A larger sample will enable us to discriminate between $w_{dg}/w_{gg} < 1$ or $w_{dg}/w_{gg} > 1$, i.e. to test whether DLA halos are more or less massive than LBG halos.

Subject headings: cosmology: observations — galaxies: evolution — galaxies: high-redshift — quasars: absorption lines — quasars: individual (APM08279+5255, PC1233+4752, J0124+0044)

1. INTRODUCTION

QSO absorption lines, including damped Ly- α absorbers (DLAs), and Lyman break galaxies (LBGs) are currently our two major sources of information on high redshift galaxies. After more than two decades of study, the exact nature and detailed characteristics of damped absorbers remain unexplained. Here, we seek to constrain the properties of DLA halos using LBGs (Steidel et al., 1999) as tracers of large scale structure.

DLAs contain the largest reservoir of neutral hydrogen (HI) at high redshifts (e.g. Lanzetta et al., 1991; Lanzetta, Wolfe, & Turnshek, 1995; Ellison et al., 2001). They contain more neutral HI than all the absorbers in the Ly-alpha forest combined. Mor-

ever, the amount of HI in DLAs at high redshifts corresponds to the amount of HI in stars today at $z = 0$: Ellison et al. (2001) find $\Omega_{HI}(z = 3) = 10^{-2.6}$, while Bell et al. (2003) measure $\Omega_*(z = 0) = 10^{-2.56}$ (both numbers are for $h = 0.65$). These facts led Wolfe et al. (1986) to put forward the ‘disk hypothesis’, namely that DLAs are large thick gaseous disk galaxies. Despite the numerous observations directed at DLAs in the past decade (e.g. imaging studies such as Møller & Warren, 1993; Lowenthal et al., 1995; Steidel et al., 1994, 1995; Le Brun et al., 1997; Bunker et al., 1999; Fynbo, Møller, & Warren, 1999; Kulkarni et al., 2000; Pettini et al., 2000; Rao & Turnshek, 2000; Bouché et al., 2001; Møller et al., 2002), this hypothesis has been debated and the role of DLAs in galaxy

¹Visiting Astronomer, Kitt Peak National Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA), Inc., under cooperative agreement with the National Science Foundation.

formation is still not understood (see the more exhaustive summary of Pettini, 2003).

Hints to the nature of DLAs are given by numerical (Katz et al., 1996; Haehnelt, Steinmetz, & Rauch, 2000; Gardner et al., 2001; Nagamine, Springel, & Hernquist, 2003b) and semi-analytical (Kauffmann, 1996; Mo, Mao, & White, 1999; Okoshi et al., 2003) simulations of galaxy formation at high-redshifts. All, these simulations indicate that DLAs are in majority faint (sub- L^*) in small dark matter halos $V_c \ll 100\text{km s}^{-1}$. Based on cross-section arguments, Fynbo, Møller, & Warren (1999), Møller et al. (2002), and Schaye (2001) arrived to the same conclusions. From the chemical evolution point of view, (Matteucci et al., 1997; Jimenez, Bowen, & Matteucci, 1999; Boissier, Péroux, & Pettini, 2003) argued that DLAs are caused by gas rich low surface brightness dwarf galaxies, as seen locally in at least one case (Bowen, Tripp, & Jenkins, 2001).

However, DLAs show asymmetric profiles of their high ionization species (Prochaska & Wolfe, 1997; Ledoux et al., 1998). This has been used to argue that DLAs are, in fact, due to thick massive rotating disks (Wolfe et al., 1986, 1995; Prochaska & Wolfe, 1997). But others, e.g. Maller et al. (2000), McDonald & Miralda-Escudé (1999) and Haehnelt, Steinmetz, & Rauch (2000), showed that a large range of morphologies can reproduce the observed kinematics: DLAs can arise from the combined effect of a massive central galaxy and a number of smaller satellites or filaments. In fact, cold gas accretion along filaments could be an important mechanism, especially at high redshifts (Keres et al., 2004).

Whether or not DLAs are indeed massive will lead to different clustering properties of the galaxies around them. In hierarchical galaxy formation models (e.g. Mo & White, 1996, 2002), this clustering yields a measurement of the dark matter halo mass associated with DLAs relative to that of the galaxies used as tracers of the large scale structure. In particular, if the galaxies are less (more) correlated with the DLAs than with themselves, this will imply that the halos of DLAs are less (more) massive than those of the galaxies. Here, we use $z \simeq 3$ LBGs (Steidel et al., 1999) as tracers of the large scale structure.

In analyses similar to that presented here, Gawiser et al. (2001) found no clustering of galaxies around one single DLA at $z = 4$ towards BR 0951-04, and Adelberger et al. (2003) found a lack of galaxies near four DLAs (they found two within a cylinder of radius of $5.7h^{-1}\text{Mpc}$ and depth $W_z < 0.025$ whereas ~ 6 were expected if the cross-correlation is the same as the galaxy auto-correlation). They argued that

this is evidence that DLAs and LBGs “do not reside in the same parts of the universe”. It is important to note that both of these studies were not sensitive to scales larger than $5h^{-1}\text{Mpc}$ because of the small field of view available.

Other studies, however, have pointed to an over-density of galaxies near DLAs. Wolfe (1993) combined several studies of Ly α emitters around DLAs and found evidence for a correlation between emitters and DLAs at a mean redshift $\bar{z} = 2.6$, significant at $> 99\%$. Francis & Hewett (1993) reported the discovery of super clustering of sub-DLAs at $z \sim 2.4$ and $z \sim 2.9$: a total of four HI clouds are seen towards a QSO pair separated by $8'$, each being at the same velocity. Recent results from narrow-band imaging of the Francis & Hewett field shows that spectroscopically confirmed Ly α emitters are clustered at the redshift of the strongest HI cloud at $z = 2.9$ ($\log N_{\text{HI}} = 20.9$) towards Q2138-4427 (Fynbo et al., 2003). Roche et al. (2000) identified eight Lyman-alpha emitting galaxies near the DLA at $z = 2.3$ towards PHL 957 in addition to the previously discovered Coup Fourré galaxy Lowenthal et al. (1991), implying the presence of a group, filament, or proto-cluster associated with the DLA. Other evidence of clustering include the work of Ellison et al. (2001) and d’Odorico, Petitjean, & Cristiani (2002). Ellison et al. (2001) found that the DLA at $z_{\text{abs}} = 3.37$ towards Q0201+1120 is part of a concentration of matter that includes at least four galaxies (including the DLA) over transverse scales greater than $5h^{-1}\text{Mpc}$. d’Odorico, Petitjean, & Cristiani (2002) showed that out of ten DLAs in QSO pairs, five are matching systems within 1000km s^{-1} . They concluded that this result indicates a highly significant over-density of strong absorption systems over separation lengths from ~ 1 to $8h^{-1}\text{Mpc}$.

Despite numerous attempts to reproduce the properties of DLAs (e.g. Katz et al., 1996; McDonald & Miralda-Escudé, 1999; Mo, Mao, & White, 1999; Gardner et al., 2001; Nagamine, Springel, & Hernquist, 2003b; Okoshi et al., 2003) in numerical simulations or semi-analytical models, there are no predictions of the clustering of galaxies around DLAs. In Bouché et al. (2004, in preparation), we discuss the cross-correlation of LBGs around DLAs in Smoothed Particle Hydrodynamical (SPH) numerical simulations. There, we find that the DLA-LBG cross-correlation is weaker than the LBG auto-correlation by approximately 30%, indicating that DLAs reside in less massive dark matter halos than those of LBGs.

In Bouché & Lowenthal (2003), we presented a preliminary measurement of clustering of LBGs around a DLA at $z \sim 3$ towards APM 08279+5255, based

on part of this data set. We detected an overdensity of $\Sigma/\Sigma_g \simeq 3$ at the 95% level on scales $2.5 < r_\theta < 5$ Mpc, implying that at least some DLAs reside in dense environments. In this paper, we expand on that study and present results on the DLA-LBG cross-correlation in our full survey of three fields.

In section 2, we present the imaging data used here, our completeness limits and pre-selection of high-redshift galaxy candidates. We used photometric redshifts to select our $z = 3$ LBG candidates as discussed in section 3. Our clustering analysis is shown in section 4. The results on the DLA-LBG cross-correlation are shown in section 5, and discussed in section 6. Section 7 contains our conclusions.

Throughout this paper, we adopt $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$ and $H_o = 100h$ km s $^{-1}$ Mpc $^{-1}$; thus, at $z \sim 3$, 1'' corresponds to $\sim 21.5h^{-1}$ kpc and 1' to $\sim 1.29h^{-1}$ Mpc, both *co-moving*. At that redshift, $H(z) \sim 4.46H_o$, so $\delta z = 0.1$ corresponds to $67h^{-1}$ Mpc in co-moving coordinates ².

2. THE DATA

2.1. The DLA fields

Given the allocated telescope night, we selected three fields for the presence of a DLA at $z \sim 3$ and with the additional constraint that the QSO must be at a higher redshift than that of the DLA, i.e. $z_{\text{abs}} \ll z_{\text{QSO}}$. The redshift $z \sim 3$ is ideal for selecting LBGs efficiently using standard photometric filters.

Our results on the QSO APM 08279+5255 field were presented in Bouché & Lowenthal (2003). The other QSOs in our survey are PC 1233+4752 ($z_{\text{em}} = 4.447$), with a DLA at $z_{\text{abs}} = 3.499$ (White, Kinney, & Becker, 1993); and SDSS J0124+0044 ($z_{\text{em}} = 3.840$), with a DLA at $z_{\text{abs}} = 3.077$ (C. Péroux, 2003, private communication). The individual QSO and DLA properties are listed in Table 1.

Although our DLAs do not meet the column density threshold of $\log N_{\text{HI}} > 20.3$ often quoted, the latter is arbitrary and based on resolution threshold of previous-generation instruments. Furthermore, the metallicities and the HI properties of ‘sub-DLAs’ are not different to the ‘strict-DLA’ population (e.g. Péroux et al., 2003). For the purpose of this study, the hydrogen column densities are $\log N_{\text{HI}} \geq 20.0$, which ensure that the absorption is damped and that the gas is neutral.

2.2. Observations

The observations were carried out with the MOSAIC camera at the Kitt Peak National Observatory 4-m telescope on UT 2000 February 7 and 8 (run I), and on UT 2001 September 23–26 (run II). Run II was photometric, some cirrus were present during run I. The seeing was $0.9\text{--}1''.5$ for both runs.

The wide field imager MOSAIC contains eight 2k×4k thinned SITe CCD. With $0''.258$ per pixel, it has a field of view of $35'$ on a side. The readout noise is $\sim 6e^-$ pix $^{-1}$, the dark current is negligible ($\sim 5e^-$ hr $^{-1}$), and the average gain is $3e^-$ ADU $^{-1}$. Each CCD has been thinned for detecting U-band photons.

We imaged our three fields through four broad-band filters—U (Stromgren) and BV&I (Harris set) (see Fig. 1)—using a standard dither pattern (five pointings) to remove cosmic rays and detector defects. The total integration time for each field was typically 4hr(U), 1hr(B & V) and 4hr(I); the observations are summarized in Table 2. In addition, we observed several Landolt (1992) standard star fields through each filter.

2.3. Data reduction

The data were reduced with the package MSCRED (v4.1) within IRAF³(v2.11.3), following the reductions guidelines of the NOAO Deep Wide-Field Survey (Jannuzi & Greer, 2000). This package was specifically designed to reduce MOSAIC data. The reduction process contains more steps than typical optical observations and is detailed here.

We first performed the overscan level subtraction. We then corrected for the small correction ($<0.5\%$) due to cross talk between adjacent chips. For each night’s data, we removed an averaged zero frame, or bias frames, from the science images. The thinned MOSAIC chips required no dark correction since the dark current was only $5e^-$ per hour.

Flat-fielding is critical to achieve precise photometry. However, we had to deal with two non-traditional complications: (i) the MOSAIC instrument at the 4-m suffers from a ghost image of the pupil in all bands due to reflections in the optics of the camera that needs to be removed; and (ii) dome flat-fields may match the night sky to only 1 or 2% (usually larger). Thus, sky-flatfielding in addition to dome-flatfielding was necessary and since the pupil image is an additive light effect, it has to be removed from the flat-fields first.

In the dome flats, we removed the pupil image by fitting an axially symmetric pattern to the data them-

²Using the latest cosmological parameters from WMAP ($\Omega_M = 0.268$, $\Omega_\Lambda = 0.728$, $h = 0.71$) changes these numbers by $\sim 3\%$.

³IRAF is distributed by the National Optical Astronomical Observatories, which are operated by AURA, Inc. under contract to the NSF

selves with the task MSCPUPIL. We then flat-fielded all the science frames with pupil-free dome-flats.

At this point, the pupil image was still present in the data. It is more difficult to remove the pupil from individual science images than from the dome-flats because the pupil pattern is much fainter, and the pupil image is mixed with all the faint and bright sources in the data. A simple threshold scheme to remove the objects is not feasible since the pupil is still present and the data poorly flat-fielded. Thus, to remove the pupil from the science images and to make the sky-flats, we had to extract the pupil image from the science images themselves through the following iterative steps: (i) we created a sky-flat (ver. 1) from the average of the science frames with a median rejection; (ii) we extracted the pupil image from the sky-flat (ver. 1) using the task MSCPUPIL (parameter ‘type’ set to ‘data’) and removed it from the science frames with RMPUPIL to produce a ‘first-pass’ pupil-free data; (iii) we created another sky-flat (ver. 2) using the pupil-free data and applied it to the data. We found that low-level light from bright stars was creating signal in the sky-flat even if strong min-max rejection was used. To solve this problem, we created object masks by using a 2σ threshold on each of the eight CCDs and masked out large areas around the brightest objects. Steps (i) through (iii) were repeated using the masks and the final sky-flat was normalized and applied to all the frames. For the I-band only, we removed fringing using the procedure in (Jannuzi & Greer, 2000) before applying the final sky-flat.

Cosmic ray removal was done using the task XZAP from the package DIMSUM and customized routines. Bad-pixel masks including the cosmic rays and bleeding regions were constructed.

De-projecting the 8 CCDs on a single image is a two step process and requires very good astrometry. First, using the coordinates of several hundred USNO stars, we interactively derived astrometric solutions ($\text{RMS} \leq 0.5''$) with MSCCMATCH for each dithered exposure. Then, we mapped the eight CCDs onto a single image by rebinning the pixels to a tangent-plane projection, thus producing pixels of constant angular size, with the task MSCIMAGE.⁴ This process matches the World Coordinate Solution (WCS) solution of all bands to a common reference position.

The individual flat-fielded, astrometrically calibrated images with a uniform zero-point were averaged with an average sigma-clipping rejection to produce the final stacked images. The scaling of each individual dithered image was performed interactively on ~ 300 astrometric calibration stars common to all

images.

For the I-band of run II, we were not able to achieve a satisfactory sky-flatfielding (residuals $\sim 1\%$). In order to correct for this, we applied a median filtering to a block-averaged image of the stacked frame and applied the normalized results to the image.

Even though all bands were deprojected to a tangent-plane solution using the same position and orientation on the sky, the relative pixel positions of objects in the different bands were not exactly identical because of dithering and effects such as flexure of the telescope, and optical distortions due to filters. Since it is important to have identical pixel positions for the photometry, we had to register and rotate slightly (~ 0.003 deg) each image with respect to the reference band (U) using ~ 25 stellar objects with high S/N in all bands throughout the field of view.

The rms in the relative astrometry of ~ 150 stellar objects with $20 < m_I < 22.5$ and $m_U < 23.5$ mag is typically ~ 0.4 pix throughout the field.

2.4. Calibration

The standard star frames, which contained ~ 150 Landolt (1992) standard stars observed through each filter and airmasses $1.0 < X < 2.5$, were reduced the same way, including the rebinning of the pixels to a tangent-plane projection. For run I, because only a handful of standard stars were available, we tied the photometric solution of run I against run II on the APM 08279+5255 field, which was re-observed in run II.

Total fluxes of the standard stars were measured by fitting a Moffat (1969) profile (similar to the $7''$ -aperture flux used by Landolt 1992 at the 0.015 mag level). Table 3 presents the best fit (computed using a linear Single Value Decomposition algorithm) to the photometric equation (e.g. Harris 1981):

$$m^{obs} = -2.5 \log \frac{C}{t_{exp}} + ZP + \alpha X + \beta (\text{color}), \quad (1)$$

where C is the number of counts, t_{exp} is the effective exposure time, ZP the photometric zeropoint, α the airmass coefficient and β the color term, which accounts for variations in the effective wavelength sampled by the filter for stars with different colors. We achieved a flux limit of typically (m_{AB}) 25.5 mag (3σ) in a $2 \times \text{FWHM}$ diameter aperture (see Table 4 for all the flux limits).

2.5. Source detection

All sources were detected in the I-band using the SExtractor (v2.1.6) software (Bertin & Arnouts,

⁴The parameter ‘fluxconserve’ was set to ‘no’ because the science exposures have been flat-fielded with flats equally distorted to yield a constant sky per pixel.

1996). This algorithm performs source detection and photometry (see Simard et al., 2002, for a summary of this package). We optimized the configuration parameters to ensure the faintest sources were detected and to optimize our completeness. We used the local background estimated in a 24-pixel wide annulus. The images were convolved with a 2 pixel FWHM Gaussian kernel before source detection. The detection threshold was set to 1.5 sigma with a minimum area of 5 pixels. Bad-pixel masks are used as flag images. SExtractor is able to perform deblending of close objects. The number of deblending sub-thresholds was set to 32 pixels, and through experimentation, the minimum contrast parameter was set to 0.0001. Our catalog contains approximately 40,000 objects per field, 30,000 of which have $I > 22.5\text{mag}$.

2.6. Photometry

For each object, we measure the color in a $2 \times \text{FWHM}$ diameter aperture, where we took seeing variations on different bands into account: the color in two bands, e.g. $(U - B) = m_B(2 \times \text{FWHM}_B) - m_U(2 \times \text{FWHM}_U)$, and similarly for other colors. Although this procedure is strictly valid only for star-like objects, it has been shown to be a good approximation for faint and unresolved galaxies (Smail et al., 1995). Indeed, from *Hubble Space Telescope* studies, the half-light radius of LBGs is, on average, $0.4''$ (Lowenthal et al., 1997), much less than our seeing. For sources that were not detected in one band (i.e. $\text{flux} < 1\sigma$), the magnitude in that band is set to the 1σ flux limit in a $2 \times \text{FWHM}$ diameter aperture and no color term is computed (Equation 1).

In the remainder of this paper, we use aperture magnitudes. To convert those to total magnitudes, we estimate the total magnitude correction for star-like objects to be $m_I(\text{tot}) = m_I(2 \times \text{FWHM}) - 0.35$ in the I-band, calibrated by adding simulated stars with known total flux into our images and measuring the recovered flux in the chosen aperture.

In addition, each object in our catalogs was corrected for Galactic extinction by adopting $E(B - V)$ values taken from the maps of Schlegel et al. (1998) assuming an $R_V = 3.1$ extinction curve.

2.7. Completeness

In order to estimate our completeness, we added to our images fake stellar objects with MOFFAT profiles that matched the image point spread function (PSF). Fluxes were measured with the same aperture. We find that we are 50% complete up to $m_I \simeq 24.35\text{mag}$. The exact values for each field are shown in Table 4. Using the transformation $I_{\text{AB}} = m_I + 0.47$, this corresponds to $I_{\text{AB}} \simeq 24.8\text{mag}$ ($\mathcal{R}_{\text{AB}} \sim 25\text{mag}$, Steidel et al. 1993) and to $0.67L^*$, where $m_R^* \simeq 24.5$ for galaxies

at $z \simeq 3$ (Steidel et al., 1999). Our 90% completeness level is $I_{\text{AB}} \simeq 24.4\text{mag}$. Thus, we reached a depth sufficient to ensure that we sample well L^* galaxies at $z = 3$.

2.8. Selecting Lyman break galaxy candidates

From our catalog of $\sim 40,000$ objects, we rejected objects close to the field edges and objects with a FWHM_IMAGE less than 0.85% the FWHM for stellar objects, presumably cosmic rays or bad pixels. We selected a subsample (1/3) using the following color cuts

$$\begin{aligned} (U - B)_{\text{AB}} &> 0.3, (B - I)_{\text{AB}} < 3.7, \\ (U - B)_{\text{AB}} &> 0.3 + 0.55 \cdot [(B - I)_{\text{AB}} - 2], \end{aligned} \quad (2)$$

and with $22.9 < I_{\text{AB}} \leq 24.8\text{mag}$. This removed most $z \leq 1$ objects and reduced the amount of computing for the photometric redshifts, which is explained in the next section. $\sim 10,000$ objects per field met the criteria shown in Eq. 2.

2.9. Stellar contamination

We did not perform any star-galaxy separation since beyond $I \simeq 23\text{mag}$, number counts are dominated by the faint galaxies at those magnitudes. However, we determined empirically the stellar contamination at faint magnitudes the following way. We first select point sources, i.e. those with FWHM of the source $< 1.15 \times \text{FWHM}$ of seeing, up to $I < 23.5\text{mag}$ and that fall in our color selection box (Eq. 2). We then select all the LBG candidates with $22.5 < m_I < 24.35\text{mag}$ that have a probability to be at $z = 3 \pm 0.25$ greater than 50%, from our photometric redshifts (see below). This subset contains some of the point sources between $m_I = 22.5\text{mag}$ and $m_I = 23.5\text{mag}$, namely those with colors consistent with being at $z \simeq 3$. Finally, we compare the number counts of the LBG candidates with that of the point sources that have $z \sim 3$ colors and extrapolate the counts at $m_I > 23.5$ to $m_I = 24.35\text{mag}$, assuming the counts to be constant at $m_I > 23.5\text{mag}$ (a conservative assumption given that the star counts almost certainly decreases from $m_I = 22.5\text{mag}$ to $m_I = 23.5\text{mag}$). We found the stellar contamination to be $< 7\%$, $< 15\%$ and $< 13\%$ for the APM 08279+5255, the PC 1233+4752, and the J0124+0044 field, respectively.

3. SAMPLE SELECTION & PROPERTIES

Following our rough color-cuts, we used photometric redshift techniques to select LBG candidates in narrow redshift slices.

3.1. Photometric Redshifts

There are two approaches to photometric redshift estimations: the empirical training set method (e.g. Koo, 1985; Connolly et al., 1995) and the spectral energy distribution (SED) fitting (Lanzetta, Yahil, & Fernández-Soto, 1996; Sawicki, Lin, & Yee, 1997; Budavári et al., 2000; Fontana et al., 2000; Csabai et al., 2003). The former is an empirical relationship between colors and redshifts determined using a multi-parametric fit. The latter is based on a set of SED templates (empirical or theoretical). The two methods are comparable in their performance at $z \leq 1$; however, the training set method is not always feasible, especially at high-redshifts (see discussion in Benítez, 2000).

On the other hand, the SED fitting method works best when there is a strong feature in the SED, such as the 4000Å break, or the 912Å Lyman break. Thus, SED fitting methods were rapidly developed for the Hubble deep field (HDF) (e.g. Budavári et al., 2000; Fernandez-Soto et al., 2001) with an accuracy of typically $\Delta z \sim 0.06(1+z)$. In our case, we used the code Hyperz from Bolzonella, Miralles, & Pelló (2000), which includes intergalactic absorption due to the Ly α forest and internal extinction A_V . We updated the intergalactic absorption prescription following Massarotti et al. (2001) and we used the extinction curve of Calzetti et al. (2000) with A_V varying from 0 to 1.2.

We used the template set made of the four empirical SEDs of Coleman, Wu, & Weedman (1980), extended in the UV ($\lambda < 1400\text{Å}$) by Bolzonella, Miralles, & Pelló (2000) using the synthetic models of Bruzual & Charlot (1993) with parameters (SFR and age) that matched the spectra at $z = 0$. Note that Fernandez-Soto et al. (2001) extended the CWW templates using the power laws of Kinney et al. (1993), and Budavári et al. (2000) used the extensions of Kinney et al. (1996). The SED templates are convolved with the MOSAIC filter response curves (including the CCD response), and z_{phot} is found from the maximum of the likelihood distribution $\mathcal{L}(z)$ derived from the χ^2 distribution.

3.2. Simple tests on the HDF

We performed several tests of the technique on the sample of 150 spectroscopically confirmed galaxies at redshifts $z \leq 6$ in the Hubble deep field north (HDF-N) (Cohen et al., 2000; Fernandez-Soto et al., 2001). We experimented with 5 template sets that included the 4 CWW templates and various starburst templates from Starburst99 (Leitherer et al., 1999). Of the 18 galaxies with $2.75 < z_{\text{phot}} < 4.5$, two (11%) are outliers with $z_{\text{spec}} \simeq 1$. From that sample, we found that (i) the four CWW templates gave the lowest

scatter $\Delta_z/(1+z_{\text{spec}}) = 0.053$, both measured using the bi-weight estimator of Beers, Flynn, & Gebhardt (1990), and (ii) $\Delta_z/(1+z_{\text{spec}})$ is not improved using near-IR photometry. Thus, near-IR observations are not required for photometric redshifts at $z \simeq 3$.

3.3. Using priors

SED fitting methods give the most likely redshift given the observed set of colors. However, information such as size, or flux, can be included in photometric redshift techniques using Bayesian probabilities (following Benítez, 2000). We coupled the SED fitting scheme with the prior likelihood distribution for a galaxy of magnitude m_I parametrized by Benítez (2000) in the following way: the product *prior* \times *likelihood* is decomposed over the SED types T :

$$P(z) = \sum_T p_T(z|m_I) \cdot \mathcal{L}_T(z), \quad (3)$$

where $p_T(z|m_I)$ is the prior probability given the galaxy magnitude m_I , and $\mathcal{L}_T(z)$ is the probability of observing the galaxy colors if the galaxy is at redshift z and has a type T . The photometric redshift z_{phot} is taken from the maximum of the $P(z)$ distribution, and the errors, σ_z , are computed from the FWHM of $P(z)$ divided by 2.35. Redshifts with large σ_z may be unreliable. A good estimator of reliability is the following (Benítez, 2000):

$$P_{\Delta z} \equiv P(|z - z_{\text{phot}}| < 0.2 \times (1 + z_{\text{phot}})), \quad (4)$$

which estimates the ‘goodness’ of a photometric redshift z_{phot} using Eq. 3, and also has the useful feature to pick likely outliers (Benítez, 2000). The factor 0.2 is arbitrary, but since the rms of photometric redshifts σ_z is $\sim 0.05(1+z)$, this factor corresponds to approximately $4 \times \sigma_z$. At $z < 6$, the overall rms of Δ_z is 0.11, and $\Delta_z/(1+z_{\text{spec}}) = 0.06$, similar to 0.059 found by Benítez (2000).

3.4. Photometric redshift distributions

Fig. 2 shows the redshift distribution of the three fields. The dotted histogram shows the photometric redshift distribution using the CWW templates with no priors. The continuous histogram shows the photometric redshift distribution using the priors. Fig. 2 shows that using the priors has the effect of removing galaxies with $z_{\text{phot}} \simeq 2$ that are likely at lower redshifts. As expected, the distribution of galaxies at $z \sim 3$ is not affected much. This is due to the fact that this method is sensitive to the shape of the SED, which has a strong break between the U and B filters at that redshift.

We used the photometric redshift z_{phot} of our galaxies to determine their absolute magnitude $M_{I,\text{rest}}$. The K -correction, which for galaxies at $z \sim 3$ corresponds to the extrapolation of their intrinsic flux at $\lambda_{\text{rest}} \sim 8000\text{\AA}$ from their observed flux at $\lambda_{\text{rest}} \sim 2000\text{\AA}$, was computed using a weighted sum on each SED. Each template was weighted by the prior probability $p_T(z|m_I) \cdot \mathcal{L}_T(z)$ since the best-fitted SED was a combination of the spectral types T (see Eq. 3). At redshift $z \sim 3$, the K -correction is, however, small: it is typically ~ 0.2 for blue SEDs, such as the Irr template of Coleman, Wu, & Weedman (1980).

Fig. 3 shows the absolute magnitude M_I as a function of z . Each dot represent one galaxy in our fields. The two continuous lines show our magnitude cuts and were computed using an Irr SED. The galaxies at redshift $z \sim 3$ are, as expected, between the two lines and near our completeness limit, which provides a consistency check of the photometric redshift technique.

Almost all points that are outside the range allowed by the continuous lines in Fig. 3 are between the dotted lines, which represent the magnitude range for an E/S0 template, and thus are best fitted by the E/S0 type. This type has a strong break at 4000\AA that creates a large K -correction of four magnitudes and makes these objects too luminous for their apparent magnitude. The fitted SEDs are likely to be wrong. At that redshift, $z_{\text{phot}} \sim 2$, IR photometry is needed to better constrain the SEDs.

3.5. Selecting reliable galaxies in slices

From the subsample described in section 2.8, we selected ~ 100 LBGs that are in a redshift slice centered on the DLA. Specifically, they were chosen to have a high probability of being at the redshift of the DLA, z_{abs} , i.e.

$$P(z_{\text{abs}} \pm W_z/2) \equiv P_{\text{DLA}} > 0.5, \quad (5)$$

where W_z is the redshift slice width. We also defined two additional redshift slices shifted by $+0.15$ or -0.15 from z_{abs} , P_+ and P_- , respectively. We choose a redshift width of $W_z = 0.15$ because, as discussed in Bouché & Lowenthal (2003), it produces the largest sample in the smallest redshift slice, given the rms of the photometric redshifts. At the end of § 6, we show that a different choice of W_z does not change the results.

More importantly, this criterion (Eq. 5) corresponds to high quality photometric redshifts as illustrated in Fig. 4 for the APM 08279+5255 field. The left panels of Fig. 4 show the probability distribution, and the right panels show the probability distribution as a function of $P_{\Delta z}$ defined in Eq. 4, both for the three different redshift slices. Fig. 4(a), (b), and (c)

show P_- , P_{DLA} and P_+ , respectively. The dots represent galaxies detected in all four bands, UBV&I. The filled squares indicate objects that are not detected in the U band. Smoothing the distributions using a Gaussian kernel (scaled to the peak) produced the continuous lines in Fig. 4. The dotted line shows the minimum threshold (> 0.5) used in selecting LBG candidates in each of the slices. From the right panels, galaxies that have a high probability of being in a redshift slice also have reliable photometric redshifts, indicated by the fact that $P_{\Delta z} \geq 0.9$. For each redshift slice, the number of galaxies that met the threshold is shown in Table 5.

Fig. 5 shows the $x - y$ positions of our LBG candidates that met the criterion Eq. 5. The square regions show the masks used to cover bright stars and defects such as streaks.

4. CLUSTERING ANALYSIS

In the current hierarchical theory of galaxy formation (e.g. see Longair, 1998, and reference therein), small quantum fluctuations that were stretched out to cosmological scales by inflation grew (mainly linearly) during the radiation-dominated era, till the present. The initial power spectrum ($P(k) \propto k^n$), which characterizes these fluctuations in Fourier space, is nearly scale invariant (i.e. $n \simeq 1$) on all scales. Initially, all scales grew at the same rate. Small scales entered the horizon before the universe became matter-dominated. During that time their growth was suppressed. The resulting power spectrum $P_E(k)$ has $n \simeq 1(-3)$ on large (small) scales. These dark-matter fluctuations formed deep gravitational potentials in which galaxies and galaxy clusters formed. When the density contrast reached $\delta\rho/\rho \sim 1$, the fluctuations grew non-linearly until $\Delta\rho/\rho \sim 200$.

Since only gravity is driving this build-up of matter, massive galaxies are more likely to be found in high-density regions, whereas low-mass galaxies are more uniformly distributed. This produces an enhancement of the clustering of massive galaxies. Therefore, the clustering properties of galaxies probe their dark-matter mass distribution. The auto-correlation $\xi(r)$ is a natural tool to study clustering in this context, since ξ_{DM} is the Fourier transform of the evolved power spectrum $P_E(k)$ (e.g. Peacock, 1999). The galaxy correlation ξ_{gg} is related to the dark-matter auto-correlation ξ_{DM} via the bias b . At a given redshift,

$$\xi_{gg}(r) = b^2(M)\xi_{\text{DM}}(r). \quad (6)$$

This bias can be computed in the Press-Schechter formalism extended by Mo & White (2002) and reference therein.

Similarly to the galaxy auto-correlation ξ_{gg} , one can define the cross-correlation ξ_{dg} between DLAs

and LBGs from the conditional probability of finding a galaxy in a volume dV at a distance $r = |\mathbf{r}_2 - \mathbf{r}_1|$, given that there is a DLA at \mathbf{r}_1 :

$$P(LBG|DLA) = n_u(1 + \xi_{dg}(r))dV_2, \quad (7)$$

where n_u is the unconditional background galaxy density. Thus, the number of neighbor galaxies in a cell of volume ΔV is given by $N_p = \bar{N}(1 + \bar{\xi}_{dg}(r))$, where $\bar{N} = n_g \Delta V$ and $\bar{\xi}$ is the cross-correlation averaged over the volume ΔV . This estimator of ξ requires an estimate of the unconditional background galaxy density n_g . There are two ways to quantify n_g in case of the cross-correlation: one way is to use galaxies spatially far from the DLA as in Bouché & Lowenthal (2003); the other way is to use the entire galaxy catalog. Here, we used the latter because of the simplicity of this method when randomizing the line of sight (see section 5.2). Naturally, large fields will yield a better estimate of n_g .

We can extend the spatial cross-correlation ξ to angular correlation function w since the former is directly related to spatial correlation functions ξ . Namely, if the selection function $\phi(z) \simeq 1/W_z$ within $[-W_z/2, W_z/2]$ and zero otherwise, then

$$w(r) = \frac{2}{W_z} \int_0^{W_z/2} \xi(\sqrt{z^2 + r^2}) dz, \quad (8)$$

or $w(r_\theta) \simeq \frac{A}{W_z} \left(\frac{r_\theta}{r_o}\right)^\beta$ where $\beta = 1 - \gamma$ and A is a constant that can be computed analytically (e.g. Adelberger et al., 2003; Eisenstein, 2003).

We used the following estimator of the cross-correlation $w_{dg}(r)$,

$$1 + \bar{w}_{dg}(r) = \left\langle \frac{N_r N_{\text{obs}}(r)}{N_g N_{\text{rand}}(r)} \right\rangle, \quad (9)$$

where $N_{\text{obs}}(r)$ is the observed number of galaxies between $r - dr/2$ and $r + dr/2$, $N_{\text{rand}}(r)$ is the number of randomly distributed galaxies, N_g is the total number of galaxies, N_r is the total number of random galaxies, and $\langle \rangle$ denotes the average over the number of DLAs (N_{DLA}). The random catalog N_r is much larger than the galaxy catalog so that the variance of $N_{\text{rand}}(r)$ is negligible.

The errors to w in Eq. 9 are given by (e.g. Landy & Szalay, 1993)

$$\sigma_w \simeq \frac{\sigma_{\langle N_{\text{obs}} \rangle}}{\langle N_{\text{obs}} \rangle} [1 + \bar{w}_{dg}], \quad (10)$$

where $\langle N_{\text{obs}} \rangle$ is the number of neighbors in the annuli averaged over N_{DLA} . Using solely the Poisson variance of the mean $\langle N_{\text{obs}} \rangle$, i.e. $\sigma_{\langle N_{\text{obs}} \rangle} =$

$\sqrt{\langle N_{\text{obs}} \rangle / N_{\text{DLA}}}$, yields the Poisson errors for Eq. 9 (e.g. Mo, Jing & Boerner, 1992; Landy & Szalay, 1993):

$$\sigma_w \simeq \frac{1 + \bar{w}_{dg}}{\sqrt{\langle N_{\text{DLA}} \rangle \langle N_{\text{obs}} \rangle}}. \quad (11)$$

In general, the cross-correlation and the galaxy auto-correlation will increase the variance to $\langle N_{\text{obs}} \rangle$, and σ_w will be larger (see Eisenstein, 2003).

Since $\langle N_{\text{obs}} \rangle$ is proportional to the total number of galaxies, N_g , the expected rms of the cross-correlation function is

$$\sigma_w \propto \frac{1}{\sqrt{\langle N_{\text{DLA}} \rangle \langle N_g \rangle}} \quad (12)$$

as one might have expected. Thus, the noise in \bar{w}_{dg} goes as the inverse of the square root of the number of DLAs, N_{DLA} , and as the inverse of the square root of the number of galaxies N_g .

5. RESULTS

With the clustering formalism layed out in section 4, we can present our results on the DLA-LBG cross-correlation (§ 5.1) and on the computation of the errors (§ 5.2).

5.1. DLA-LBG cross-correlation from the combined fields

Fig. 6 shows the DLA-LBG cross-correlation w_{dg} computed using Eq. 9. In computing $N_{\text{obs}}(r)$ and N_{rand} in Eq. 9, we took into account the masked regions with bright stars shown in Fig. 5. The dotted line shows the auto-correlation w_{gg} of Adelberger et al. (2003) and the continuous line shows a fit to the amplitude of w_{dg} using w_{gg} as template. The fitting method is described below.

It is necessary to take into account the different selections of the different fields in performing the sum in Eq. 9 indicated by the brackets, so we must weight each field accordingly. We chose to weight each field according to its errors at each angular scale r_i . Thus, for each field l , we compute $\frac{N_{\text{obs}}}{N_{\text{rand}}} \equiv 1 + w_l(r)$ in six annuli. Then, the combined angular cross-correlation, w_{dg} is computed from the weighted mean of $w_l(r_i)$, where the weights were $\sigma_l(r_i)$ at each angular scale r_i .

The error bars shown in Fig. 6, $\sigma(r_i)$, are from the full covariance matrix to w_{dg} . The latter is found using the standard error propagation formula:

$$\text{COV}_{i,j} = \sum_{l=1}^{N_{\text{DLA}}} \frac{\partial f(r_i)}{\partial w_l(r_i)} \frac{\partial f(r_j)}{\partial w_l(r_j)} \text{COV}_l(r_i, r_j) \quad (13)$$

where N_{DLA} is the number of DLAs, $f(r)$ is the weighted mean $w_{dg}(r)$ and $\text{COV}_l(r_i, r_j)$ is the covariance matrix of the individual field l given by Eq. 14 (below).

5.2. Error computation

Because each DLA is at a slightly different redshift, each field has a different selection function, and it is necessary to take into account these differences by weighting each field accordingly. As mentioned, we chose to weight each field according to its errors $\sigma_w(r_i)$.

The errors need to be computed carefully. Several options are available. The proper way to compute the errors would be to resample the DLAs (via bootstrap techniques), but this is impractical here given the number of DLA fields at our disposal. Another way would be to bootstrap the galaxies, which would reproduce only the Poissonian errors (Eq. 11).

We used yet another method, which is to perform Monte Carlo simulations in which we randomize the position of the DLA. This takes into account the clustering variance due to the galaxy auto-correlation, but misses the variance (and co-variance) due to the cross-correlation itself (the factor $1 + w_{dg}$ in Eq. 10). However, this term will be small on scales larger than $5h^{-1}$ Mpc because $w_{dg} \ll 1$.

Thus, in each DLA field l , we computed the full covariance matrix COV_l from $n_r = 200$ randomizations of the DLA position:

$$\text{COV}_l(r_i, r_j) = \frac{1}{n_r - 1} \sum_k^{n_r} [w_k(r_i) - \bar{w}(r_i)][w_k(r_j) - \bar{w}(r_j)] \quad (14)$$

where w_k is the k th measurement of w and \bar{w} is the average of the n_r measurements of the cross-correlation. The errors $\sigma_l(r_i)$ to $w_l(r_i)$ for each field l follow.

Our errors are consistent with the errors expected from our analysis of cosmological simulations: in Bouché et al. (2004, in preparation), we conclude that with a data set of this size, we can be sensitive to the cross-correlation only on scales $5\text{-}10h^{-1}$ Mpc, which is where we see a positive cross-correlation.

5.3. The integral constraint

Because the unconditional galaxy density, n_u in Eq. 7, is estimated from the total observed galaxy density, whereas it should always be lower than the observed galaxy density, all estimates of ξ (or w) are biased low. This bias Δw , often referred to as the ‘integral constraint’, can be computed analytically (e.g. Landy & Szalay, 1993; Saslaw, 2000). For the angular cross-correlation function, it is :

$$\Delta w = \frac{1}{\Omega} \int d\Omega \hat{w}_{dg}(r_\theta), \quad (15)$$

where \hat{w}_{dg} is a model of the cross-correlation and Ω is the total survey area. Thus, Δw will be smaller with

larger fields, and we corrected w_{dg} in Fig. 6 for its bias C due to the integral constraint assuming that w_{dg} is w_{gg} from Adelberger et al. (2003).

6. DISCUSSION

We next discuss quantitatively how significant the cross-correlation shown in Fig. 6 is. We also show that our measurement of the DLA-LBG cross-correlation (i) is not reproduced by random lines of sight (at the 95% level) and (ii) is not seen in other redshift slices that do not contain the DLAs. We compare these results to past and future studies at the end of this section.

6.1. Is this result consistent with no clustering?

It is clear from Fig. 6 that we will not be able to constrain the slope of the cross-correlation, so we assumed that the cross-correlation w_{dg} has the same shape as the auto-correlation w_{gg} . That is, we fitted the cross-correlation to the model

$$\hat{w}_{dg} = a \times \hat{w}_{gg}, \quad (16)$$

where a is the unknown amplitude.

We fitted the amplitude of the cross-correlation a using the covariance $\text{COV}_{i,j}$ computed in Eq. 13, using the following χ^2 statistic:

$$\chi^2 = \sum_{i,j} [w_{dg}(r_{\theta_i}) - \hat{w}_{dg}(r_{\theta_i})] \text{COV}_{i,j}^{-1} [w_{dg}(r_{\theta_j}) - \hat{w}_{dg}(r_{\theta_j})], \quad (17)$$

where \hat{w}_{dg} is the model of the angular cross-correlation and COV^{-1} is the inverse of the covariance matrix computed using a Single Value Decomposition algorithm. The $\chi^2(a)$ distribution is shown in the small panel in Fig. 6. In performing the inversion of the covariance matrix, we rejected the eigenvalue corresponding to the last radial bin or to scales similar to the size of the field where there is no signal (see discussion in Bernstein [1994]).

We found that an amplitude of $a > 0$ was favored:

$$a = 1.62 \pm 1.32. \quad (18)$$

This measurement of the DLA-LBG cross-correlation is most significant on scales greater than $5\text{-}10h^{-1}$ Mpc. At this point, we cannot, however, conclude whether the DLA-LBG cross-correlation is stronger or weaker than the LBG-LBG auto-correlation. Taking $r_o \simeq 4h^{-1}$ Mpc and $\beta \simeq 0.6$ for the LBG auto-correlation (Porciani & Giavalisco, 2002; Adelberger et al., 2003), we find that the correlation length of the cross-correlation w_{dg} is $r_o = 5 \pm 4.5h^{-1}$ Mpc (or $r_o = 7 \pm 6.8h_{71}^{-1}$ Mpc).

6.2. *Is this result drawn from random lines of sight?*

Given the large rms to the fitted amplitude a , could our result simply be a large fluctuation of the set of possible values for random lines of sight? To test this, we chose 100 lines of sight selected at random, excluding the central $5h^{-1}$ Mpc to ensure that the new lines of sight are not correlated with the ones centered on the DLAs. We then computed the cross-correlation for these 100 random lines of sight in the redshift slices centered on the DLAs. As before, we computed the weighted mean to w_{dg} and used Eq. 13.

Fig. 7 shows the logarithm of the $\chi^2(a)$ for fixed amplitudes a for the 100 random lines of sight (filled circles). The continuous line shows the median of the distributions. The dotted and dashed lines are the 95% and 99% levels of the distributions. The median, 95% and 99% levels are found after a Gaussian kernel smoothing of the distributions using the optimum band width (Wand & Jones, 1995) (The results are not significantly changed using a fixed band width). The open square shows the location of the result of Fig. 6. Since it lies close to the 95% confidence level, this shows that the signal measured in Fig. 6 is not drawn from a random distribution of lines of sight, at the $> 95\%$ confidence.

6.3. *How about other redshift slices?*

The result of Fig. 6 should be compared with the cross-correlation when there is no DLA in the redshift bin. From our photometric redshift analysis, we selected galaxies in two other redshift slices that did NOT contain the DLA. We chose the slices that were in the foreground and in the background from the DLA, and offset by $+0.15$ or -0.15 in redshift (see Fig. 4). In each case, the χ^2 fit does not favor any clustering: the best amplitude is $a = -0.20 \pm 1.26$ and $a = -0.24 \pm 2.04$, respectively. A clustering signal in this slice would have cast a strong doubt on our results that do contain the DLA in Fig. 6. In addition, we performed the same check on another slice at redshift 3.6. The best amplitude for this slice is $a = -0.13 \pm 1.44$.

We repeated the analysis with $W_z = 0.20$ to test whether the observed clustering depends on the choice of the slice width. We found that $a = 1.45 \pm 1.35$ in this case, so we conclude that the slice width does not strongly affect the clustering signal.

6.4. *Comparison with past and future work*

Wolfe (1993) also found that Ly-emitters are strongly clustered around DLAs. In contrast, Gawiser et al. (2001) did not find evidence of clustering and the study of Adelberger et al. (2003) found a lack of galaxies near their four DLAs, within

$5.7h^{-1}$ Mpc. Since these two surveys were not sensitive to clustering on scales larger than $> 5h^{-1}$ Mpc, and ours is not sensitive to $< 3 - 5h^{-1}$ Mpc, our results are not inconsistent with theirs. The lack of galaxies on small scales could, however, be due to more local environmental effects, such as strong galactic winds from star forming galaxies.

Although simulations of DLA properties exist (e.g. Katz et al. (1996); Gardner et al. (2001); Nagamine, Springel, & Hernquist (2003b)), no prediction of the DLA-LBG cross-correlation has been published. In Bouché et al. (2004, in preparation), we use the Tree-Smoothed Particle Hydrodynamical (TreeSPH) cosmological simulations of Katz, Weinberg, & Hernquist (1996b) to measure the ‘theoretical’ DLA-LBG cross-correlation. These simulations contain 128^3 dark matter particles and as many gas or star particles. Each galaxy ($\equiv > 64$ SPH bound particles) is able to form stars. With a similar number of DLA-LBG pairs and redshift depth ($W_z = 111h^{-1}$ Mpc), we find $w_{dg} > 0$ with the same signal to noise. Furthermore, with a much larger sample of 200 simulated DLAs, we find $w_{dg} \simeq 0.75 \pm 0.1$, or $r_o \simeq 3.5h^{-1}$ Mpc. Thus, DLA halos are less massive than the halos of LBGs, which are $10^{12}M_\odot$ (Porciani & Giavalisco, 2002; Ouchi et al., 2003). Given the present sample of three DLAs, our observed constraint on w_{dg} with its uncertainty is consistent with these simulation results.

7. SUMMARY AND CONCLUSIONS

Based on deep ($\mu_{I,AB}(sky) \simeq 27.6$ mag arcsec $^{-2}$) wide-field images (0.31deg^2 or $\sim 65 \times 65h_{71}^{-1}$ Mpc co-moving at redshift $z = 3$) around three DLAs, we identify LBG candidates brighter than $I_{AB} = 24.80\text{mag}$ using photometric redshift techniques that included the I magnitude as a prior estimate in addition to the colors.

From the redshift likelihood distributions, we selected LBG galaxies within a redshift slice of width $W_z = 0.15(\simeq \sigma_z)$ centered on the redshift of the DLAs z_{abs} . Within that slice, we cross-correlated the LBGs with the position of the DLAs and found that

- the amplitude of the DLA-LBG cross-correlation w_{dg} relative to the auto-correlation w_{gg} was $w_{dg}/w_{gg} \equiv a = 1.62 \pm 1.32$, corresponding to $r_o = 5 \pm 4.5h^{-1}$ Mpc (co-moving),
- the amplitude of the DLA-LBG cross-correlation is $a > 0$, which is significant at the $> 95\%$ confidence level based on Monte Carlo simulations,
- the clustering signal was not present in three redshift slices that did not contain the DLAs.

Given the uncertainty of our results, we cannot put constraints on the halo masses of DLAs and discriminate between the large disk hypothesis (e.g Wolfe et al., 1986), and small sub- L^* hypothesis (e.g Maller et al., 2000; Haehnelt, Steinmetz, & Rauch, 2000; Møller et al., 2002). Our observation of the clustering on large scales ($> 4h^{-1}$ Mpc) is not inconsistent with previous clustering studies (Gawiser et al., 2001; Adelberger et al., 2003) since these were limited to small scales. In order to be able to directly compare these studies with our present results on scales $< 4h^{-1}$ Mpc, a larger sample of DLAs and multi-object spectroscopy of our LBG can-

didates are needed. This will enable to test whether the cross-correlation is stronger or weaker than the auto-correlation.

N.B. acknowledges a post-doc fellowship from the European Community Research and Training Network “The Physics of the Intergalactic Medium”. J. D. L. acknowledges support from NSF grant AST-0206016. We thank the anonymous referee for a careful reading of the manuscript that improved the quality of the paper. We also thank H. Mo, N. Katz and B. Ménard for helpful discussions, and J. Fynbo for reading a earlier draft.

References

- Adelberger, K. L., Steidel, C. C., Shapley A. E., & Pettini M. 2003, ApJ, 584, 45
- Beers, T. C., Flynn, K., & Gebhardt, K. 1990, AJ, 100,32
- Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. 2003, ApJS, 149, 289
- Benítez N. 2000, ApJ, 536, 571
- Bertin, E., & Arnouts, S. 1996, A&A, 117, 39
- Boissier, S., Péroux, C., & Pettini, M. 2003, MNRAS, 338, 131
- Bolzonella, M., Miralles, J-M., & Pelló, R. 2000, A&A, 363, 476
- Bouché, N., Lowenthal, J. D., Charlton, J. C., Bershad, M. A., Churchill, C. W., & Steidel, C. C. 2001, ApJ, 550, 585
- Bouché, N., & Lowenthal, J. D. 2003, ApJ, 596, 810
- Bouché, N., Gardner, J. P., N. Katz, Davé, R., Weinberg, D., & Lowenthal, J. D. 2004, in preparation
- Bowen, D. V., Tripp, T. M., & Jenkins, E. B. 2001, AJ, 121, 1456
- Bernstein, G. M. 1994, ApJ, 424, 577
- Bruzual A., G., & Charlot, S., 1993, ApJ, 405, 538
- Budavári, T., Szalay, A. S., Connolly, A. J., Csabai, I., Dickinson, M. 2000, AJ, 120, 1588
- Bunker, A. J., Warren, S. J., Clements, D. L., Williger, G. M., & Hewett, P. C. 1999, MNRAS, 309, 875
- Csabai, I., Budavári, T., Connolly, A. J., Szalay, A. S., et al 2003, ApJ, AJ, 580, 125
- Calzetti, D., Armus, L., Bohlin, R. C., Kinney, A. L., Koorneef, J., & Storchi-Bergmann, T. 2000, ApJ, 533, 682
- Carilli, C., et al. 2001, ApJ, 555, 625
- Cohen, J. G., Hogg, D. W., Blandford, R., Cowie, L. L., Hu, E., Songaila, A., Shopbell, P., & Richberg, K. 2000, AJ, 538, 29
- Coleman, G. D., Wu, C-C., & Weedman, D. W. 1980, ApJS, 43, 493
- Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., Munn, J. A. 1995, AJ, 110, 2655
- d’Odorico, V., Petitjean, P., & Cristiani, S. 2002, A&A, 390, 13

- Eisenstein, D. J. 2003, ApJ, 586, 718
- Ellison, S. L., Pettini, M., Steidel, C. C., & Shapley, A. E. 2001, ApJ, 549, 770
- Ellison, S. L., Yan, L., Hook, I. M., Pettini, M., Wall, J. V., & Shaver, P. 2001, A&A, 379, 393
- Fernández-Soto, A., Lanzetta K.M., Chen H-W., Pascarelle S. M., & Yahata, N. 2001, ApJS, 135, 41
- Fontana, A., D'Odorico, S., Poli, F., Giallongo, E., Arnouts, S., Cristiani, S., Moorwood, A., Saracco, P. 2000, AJ, 120, 2206
- Francis, P. J., & Hewett, P. C. 1993, AJ, 105, 1633
- Fynbo, J. P. U., Møller, P., & Warren, S. J. 1999, MNRAS, 305, 849
- Fynbo, J. P. U., Ledoux, C, Møller, P., Thomsen, B., & Burud, I. 2003, A&A, 407, 147
- Gardner, J. P., Katz, N., Hernquist, L., & Weinberg, D. H. 2001, ApJ, 559, 131
- Gawiser, E., Wolfe, A. M., Prochaska, J. X., Lanzetta, K. M., Yahata, N., & Quirrenbach, A. 2001, ApJ, 563, 628
- Harris, W. E., FitzGerald, M. P., & Reed, B. C. 1981, PASP, 93, 507
- Haehnelt, M. G., Steinmetz, M., & Rauch, M. 2000, ApJ, 534, 594
- Jannuzi B. T., Claver, J., & Valdes, F. 2000, available at <http://www.noao.edu/noao/noadeep/ReductionOpt/frames.htm>
- Jimenez, R., Bowen, D. V., & Matteucci, F. 1999, ApJ, 514, 83L
- Kauffmann, G. 1996, MNRAS, 281, 475
- Katz, N., Weinberg, D. H., Hernquist, L., & Miralda-Escudé, J. 1996, ApJ, 457, 57L
- Katz, N., Weinberg, D. H., Hernquist, L., 1996b, ApJS, 105, 19
- Keres, D., Katz, N., Weinberg, D. H., Davé, R., 2004, ApJ, in preparation
- Kinney, A. L., Bohlin, R. C., Calzetti, D., Panagia, N., & Wyse, R. F. G. 1993, ApJS, 86, 5
- Kinney, A. L., Calzetti, D., Bohlin, R. C., McQuade, K., Storchi-Bergmann, T., Schmitt, H. R. 1996, ApJ, 467, 38
- Koo, D. 1985, AJ, 90, 418
- Kulkarni, V. P., Hill, J. M., Schneider, G., Weymann, R. J., Storrie-Lombardi, L. J., Rieke, M. J., Thompson, R. I., & Jannuzi, B. T. 2000, ApJ, 536, 36
- Landolt, A. 1992, AJ, 104, 340
- Landy, S. D., & Szalay, A. 1993, ApJ, 412, 64
- Lanzetta, K. M., McMahon, R. G., Wolfe, A. M., Turnshek, D. A., Hazard, C., & Lu, L. 1991, ApJS, 77, 1
- Lanzetta, K. M., Wolfe, A. M., & Turnshek, D. A. 1995, ApJ, 440, 435
- Lanzetta K. M., Yahil A., & Fernández-Soto A. 1996, Nature, 381, 759
- Le Brun, V., Bergeron, J., Boissé, P., & Deharveng, J. M. 1997, A&A, 321, 733
- Ledoux, C., Petitjean, P., Bergeron, J., Wampler, E. J., & Srianand, R. 1998, A&A, 337, 51
- Leitherer, C., et al. 1999, ApJS, 123, 3

- Lewis G. F., Robb R. M., & Ibata R. A. 1999, *PASP*, 111, 1503
- Longair, M. S. 1998, “Galaxy formation”, eds Springer-Verlag
- Lowenthal, J. D., Hogan, C. J., Green, R. F., Green, R. F., Caulet, A., Woodgate, B. E., Brown, L., & Foltz, C. B. 1991, *ApJ*, 377, 73L
- Lowenthal, J. D., Hogan, C. J., Green, R. F., Woodgate, B. E., Caulet, A., Brown, L., & Bechtold, J. 1995, *ApJ*, 451, 484
- Lowenthal, J. D., Hogan, C. J., Green, R. F., Caulet, A., Woodgate, B. E., Brown, L., & Foltz, C. B. 1991, *ApJ*, 377, L73
- Lowenthal, J. D., Koo, D. C., Guzman, R., Gallego, J., Phillips, A. C., Faber, S. M., Vogt, N. P., Illingworth, G. D., & Gronwall, C. 1997 *ApJ*, 481, 673
- Maller, A. H., Prochaska, J. X., Somerville, R. S., & Primack, J. R. 2000, in *ASP Conf. Ser.* 200, eds. A. Mazure, O. Le Fèvre, and V. Le Brun, p.430
- Massarotti, M., Iovino, A., Buzzoni, A., & Valls-Gabaud, D. 2001, *A&A*, 380, 425
- Matteucci, F., Molaro, P., Vladilo, G. 1997, *A&A*, 321, 45
- McDonald, P., & Miralda-Escudé, J. 1999, *ApJ*, 519, 486
- McMahon, R. G., White, R. L., Helfand, D. J., Becker, R. H. 2002, *ApJS*, 143, 1
- Moffat, A. F. J., 1969, *A&A*, 3, 455
- Mo, H. J., Jing, Y. P., & Boerner, G. 1992, *ApJ*, 392, 452
- Mo, H. J., & White, S. D. M. 1996, *MNRAS*, 282, 347
- Mo, H. J., Mao, S., & White, S. D. M. 1999, *MNRAS*, 304, 175
- Mo, H. J., White, S. D. M. 2002, *MNRAS*, 336, 112
- Møller, P., & Warren, S. J. 1993, *A&A*, 270, 43
- Møller, P., Warren, S. J., Fall, S. M., Fynbo, J. U., & Jakobsen, P. 2002, *ApJ*, 574, 51
- Nagamine, K., Springel, V., & Hernquist, L. 2003, *MNRAS*, 348, 421
- Okoshi, K., Nagashima, M., Gouda, N., & Yoshioka, S., 2004, *ApJ*, 603, 12
- Ouchi, M. et al. 2003, *ApJ*, in press (astro-ph/0309657)
- Peacock, J. A. 1999, “Cosmological physics”, eds Cambridge University Press.
- Péroux, C., Dessauges-Zavadsky, M., D’Odorico, S., Kim, T.-S., & McMahon, R. G. 2003, *MNRAS*, 345, 480
- Petitjean, P., Aracil, B., Srianand, R., & Ibata, R. A. 2000, *A&A*, 359, 457
- Pettini, M., Ellison, S. L., Steidel, C.C., Shapley, A. E., & Bowen, D.V. 2000, *ApJ*, 532, 65
- Pettini, M. 2003, XIII Canary Islands Winter School of Astrophysics, ‘Cosmochemistry: The Melting Pot of Elements’ (astro-ph/0303272)
- Porciani, C., & Giavalisco, M. 2002, *ApJ*, 565, 24
- Prochaska, J. X., & Wolfe, A. M. 1997, *ApJ*, 487, 73
- Rao, S. M., & Turnshek, D. A. 2000, *ApJS*, 130, 1

- Roche, N., Lowenthal, J., & Woodgate, B. 2000, MNRAS, 317, 937
- Saslaw, W. C. 2000, in “The distribution of the galaxies. Gravitational clustering in cosmology”, Cambridge University Press
- Sawicki, M. J., Lin H. & Yee H. K. C. 1997, AJ, 113, 1
- Schaye, J. 2001, ApJ, 559, L1
- Schlegel, D., Finkbeiner, D., & Davis, M., ApJ, 1998, 500, 525
- Schneider D. P., Schmidt, M., & Gunn, J. E. 1991, AJ, 101, 2004
- Simard, L., et al. 2002, ApJS, ApJS, 142, 1
- Smail, I., Hogg, D. W., Yan, L., Cohen, J. G. 1995, ApJ, 449, 105
- Steidel, C. C., & Hamilton, D. 1993, AJ, 105, 2017
- Steidel, C. C., Pettini, M., Dickinson, M., & Persson, S. E. 1995, AJ, 108, 2046
- Steidel, C. C., Bowen, D. V., Blades, J. C., Dickinson, M. 1995, ApJ, 440, L45
- Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1
- Véron-Cetty, M.-P., & Véron, P. 2001, A&A, 374, 92
- Wand, M. P., & Jones, M. C. 1995, “Kernel smoothing”, eds. Chapman & Hall.
- White, R. L., Kinney, A. L., & Becker, R. H. 1993, AJ, 407, 456
- Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249
- Wolfe, A. M. 1993, ApJ, 402, 411
- Wolfe, A. M., Lanzetta, K. M., Foltz, C. B., & Chaffee, F.H. 1995, ApJ, 454, 698

TABLE 1
 PROPERTIES OF THE DLAS AND THE QSOs.

	APM 08279+525	PC 1233+4752	J0124+0044
R.A. (J2000) ..	08 ^h 31 ^m 41.6 ^s	12 ^h 35 ^m 31.1 ^s	01 ^h 24 ^m 03.8 ^s
Dec. (J2000) ..	52° 45' 17"	47° 36' 06"	00° 44' 33"
<i>l</i>	165° 45' 17"	130° 32' 12"	139° 58' 19"
<i>b</i>	36° 14' 25"	69° 17' 27"	−61° 02' 43"
A_U ^a (mag)	0.20	0.08	0.13
E_{B-V}	0.04	0.02	0.03
QSO Properties			
QSO	BAL		
z_{QSO}	3.81	4.447	3.840
m_R	15.2 ¹	20.63 ¹	17.9 ¹
Radio	$S_{20\text{cm}} = 1.3\text{mJy}$ ²	$S_{6\text{cm}} < 77\mu\text{Jy}$ ³	$S_{20\text{cm}} = 0.11\text{mJy}$ ⁴
DLA properties			
z_{abs}	2.974	3.499 ⁵	3.077 ⁷
HI W_r (Å) ..	> 4.8 ⁶	4.22 ⁵	...
$\log N_{\text{HI}}$ (cm ^{−2})	< 20.3 ⁶	20.9 ⁵	20.1 ⁷
[Fe/H]	−2.31 ⁶

^aGalactic extinction from Schlegel et al. (1998), averaged over the field.

REFERENCES.— (1) Véron-Cetty & Véron (2001); (2) McMahon et al. (2002) (3) Schneider, Schmidt, & Gunn (1991) (4) Carilli et al. (2001) (5) White, Kinney, & Becker (1993); (6) Petitjean et al. (2000); (7) Péroux, C., 2003, private communication.

TABLE 2
 SUMMARY OF THE OBSERVATIONS.

Field	U band	B band	V band	I band	UT Date of Obs.
	Total Exposure Time				
APM 08279+5255	3.75hr	35min	50min	2.08hr	Feb. 7, 8, 2000
PC 1233+4752	3.50hr	40min	50min	1.92hr	Feb. 7, 8, 2000
J0124+0044	3.72hr	47min	52min	2.08hr	Sept. 23–26, 2001

TABLE 3
PHOTOMETRIC SOLUTION.

Filter	Run I ^a		Run II	
	ZP	ZP	α	β
U	23.26 (0.08)	23.52 (0.02)	-0.421 (0.02)	0.018 (0.007)
B	25.085 (0.03)	25.26 (0.02)	-0.180 (0.01)	0.095 (0.008)
V	25.07 (0.02)	25.25 (0.02)	-0.072 (0.01)	-0.025 (0.009)
I	24.58 (0.05)	24.76 (0.02)	0.020 (0.01)	0.003 (0.011)

^aFor run I, we assumed the airmass coefficient α and the color term β to be the same as for run II.

TABLE 4
DEPTH OF THE OBSERVATIONS

Fields	Filter	Exp. /Frames (sec./#)	Airmass X^a (min-max)	FWHM (arcsec)	$SB_{lim}(1\sigma)^b$ (mag/m _{AB})	$SB_{lim}(5\sigma)^b$ (mag/m _{AB})	$m_{lim}(3\sigma)^c$ (mag/m _{AB})
APM 08279+5255	U	13500/15	1.07-1.17	1.1	27.78/28.49	26.03/26.74	25.85/26.56
	B	2100/7	1.14-1.27	1.1	28.44/28.36	26.69/26.61	26.49/26.42
	V	3000/10	1.21-1.37	1.2	28.21/28.23	26.46/26.48	26.17/26.19
	I	7590/20	1.11-1.62	1.1	27.15/27.61	25.40/25.86	25.21/25.66
PC 1233+4752	U	12600/14	1.04-1.16	1.05	27.82/28.52	26.07/26.78	25.94/26.65
	B	2400/8	1.05-1.07	1.0	28.59/28.51	26.84/26.76	26.75/26.68
	V	7590/10	1.07-1.11	0.9	28.18/28.20	26.43/26.45	26.45/26.47
	I	6900/15	1.07-1.45	1.1	27.19/27.64	25.44/25.90	25.30/25.76
J0124+0044	U	13400/16	1.19-1.51	1.5	27.88/28.59	26.13/26.84	25.60/26.31
	B	2800/7	1.24-1.42	1.5	28.69/28.61	26.94/26.87	26.44/26.37
	V	3100/7	1.46-1.90	1.4	28.33/28.35	26.58/26.60	26.14/26.16
	I	7500/20	1.21-1.80	1.1	27.67/28.13	25.93/26.38	25.75/26.21

^aThe airmass is $\frac{1}{\cos\zeta}$ where ζ is the zenith angle of the telescope.

^bLimiting surface brightness in magnitudes per square arcsecond.

^cMeasured inside a $2\times$ FWHM diameter aperture.

TABLE 5
NUMBER OF GALAXIES IN THE DIFFERENT REDSHIFT SLICES.

Field	P_-	P_{DLA}	P_+
APM 08279+5255	89	84	17
PC 1233+4752	20	70	83
J0124+0044	184	65	22

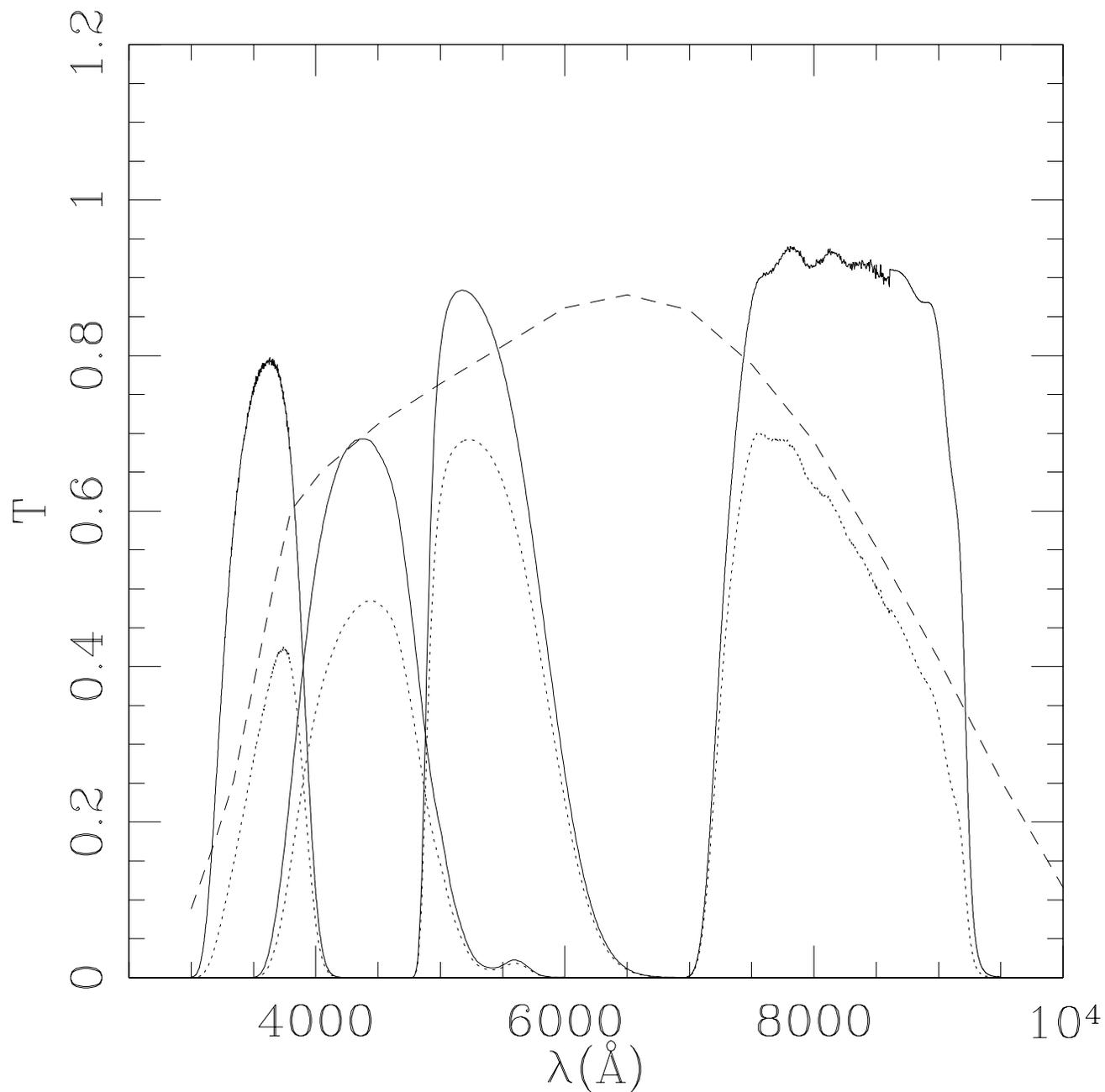


FIG. 1.— The solid lines show the transmission curves for our four filters U , B , V , and I . The dashed line shows the CCD response function. The dotted lines show the filter transmission convolved with the CCD response function.

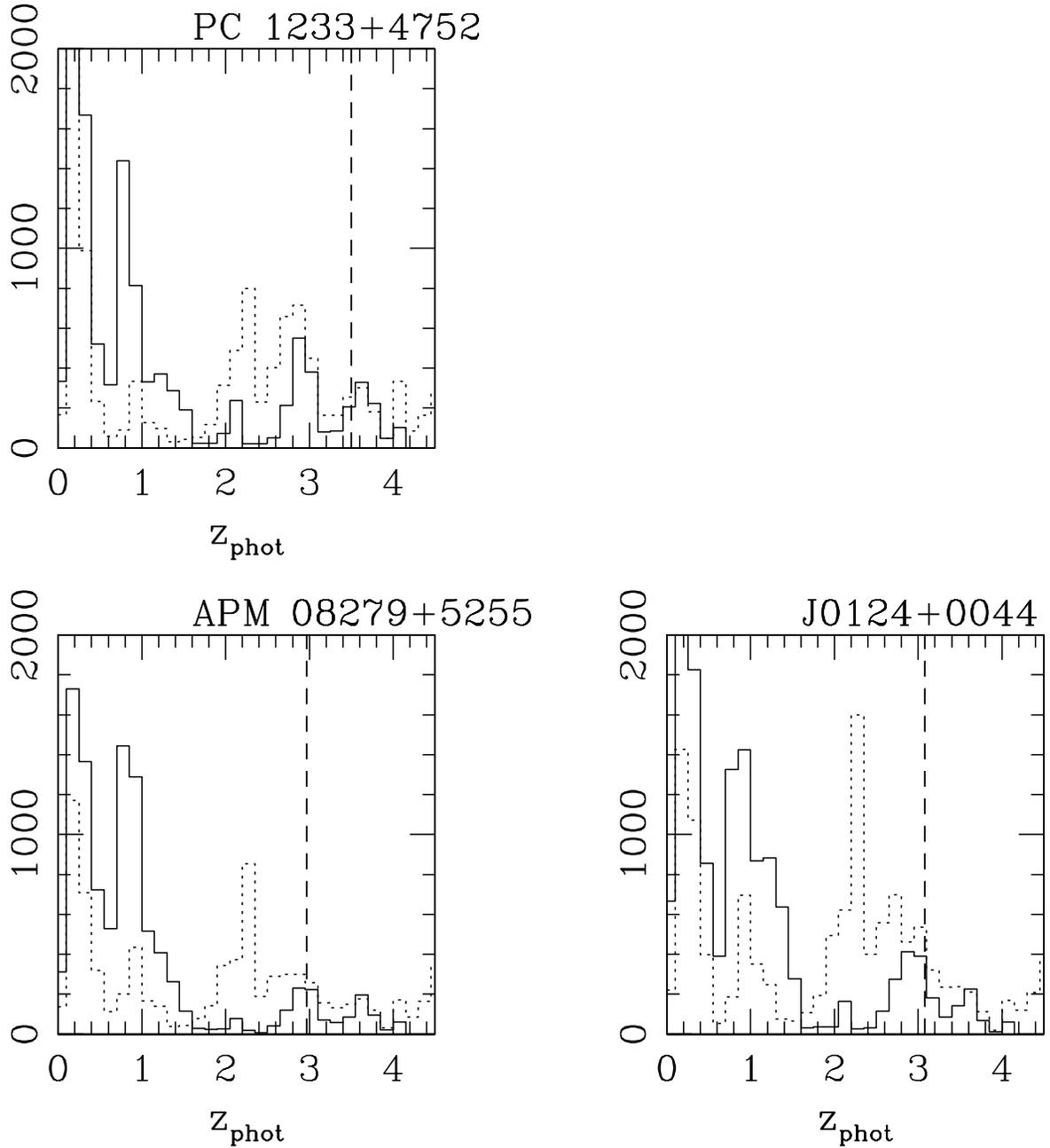


FIG. 2.— Redshift distribution for each of our fields. The dotted histogram shows the photometric redshift distribution using no priors and the template set A. The continuous histogram shows the photometric redshift distribution using the priors. Using the priors has the effect of eliminating the large number of galaxies that have been assigned $z_{\text{phot}} \simeq 2$ wrongly, but does not affect the distribution at $z \sim 3$ significantly. The vertical dashed line shows the redshift of the DLA z_{DLA} . This plot shows the effect of the priors and that our selection peaks at a redshift close to that of the DLA.

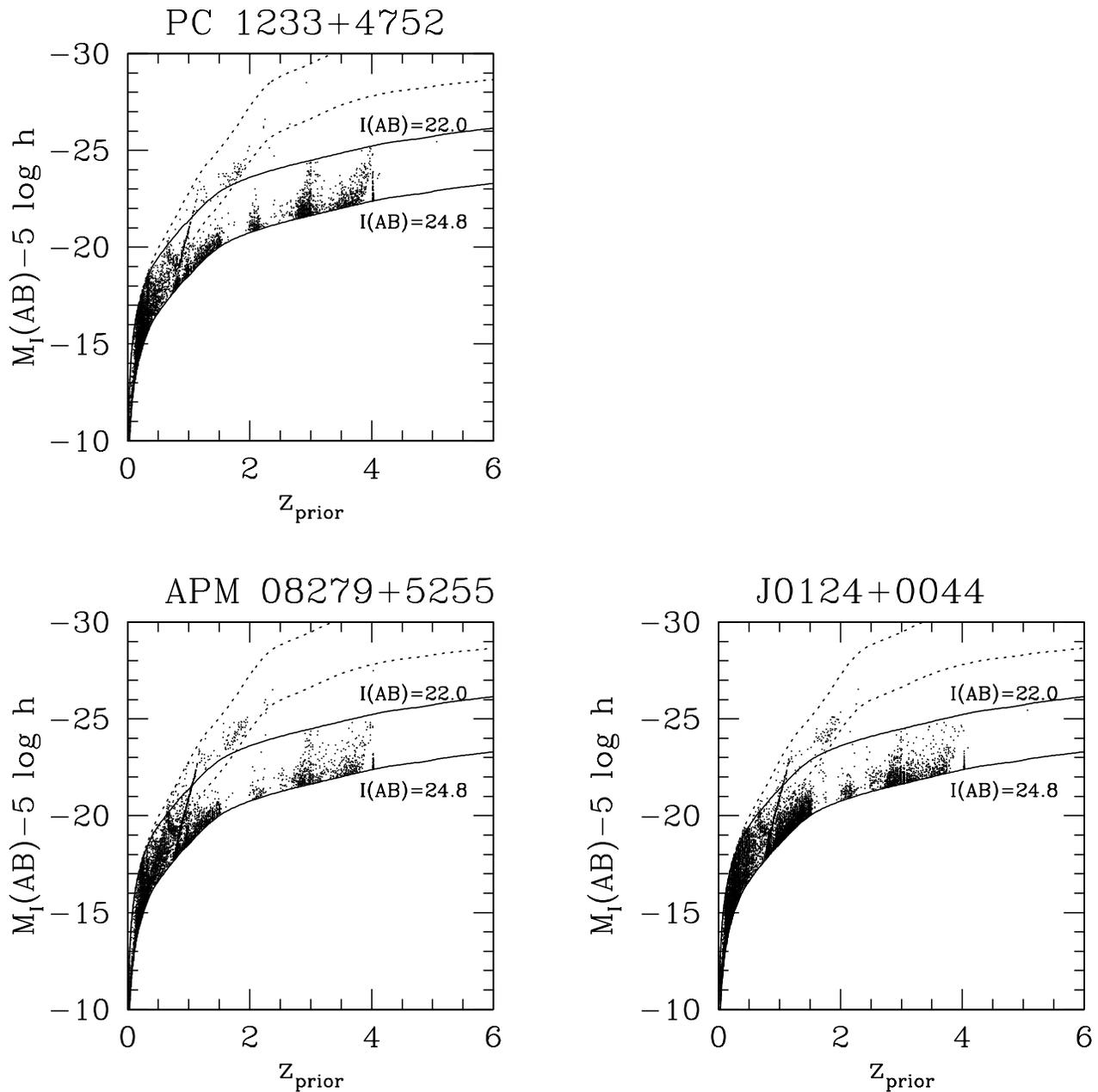


FIG. 3.— Absolute magnitude vs. photometric redshift for each field. Each dot represents one galaxy in our fields. The rest-frame absolute magnitude M_I was computed using the distance modulus and the K -correction (see text for details). The continuous lines show our magnitude selection $22 < I_{AB} < 24.8$ mag and are computed using an Irr SED. At $z \sim 3$, galaxies are between the two continuous lines and near our completeness limit as expected, which gives us more confidence in the photometric redshifts. The dotted lines show our magnitude selection for an E/S0 SED. Clearly, points that are outside the range allowed by the continuous lines are best-fitted by the E/S0 type, which has a strong break at 4000\AA and thus a large K -correction. This plot shows that the photometric redshift technique is self-consistent.

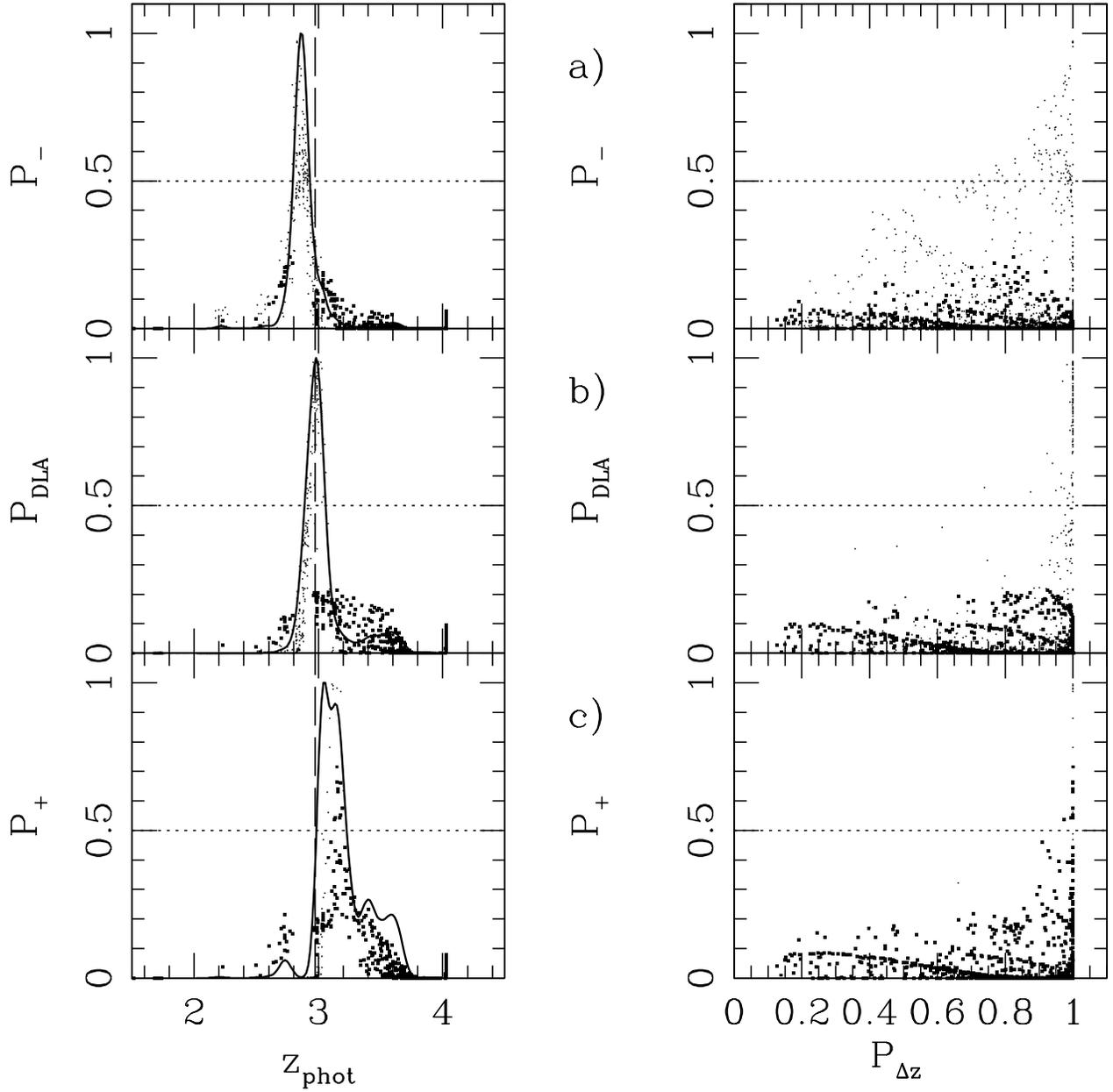


FIG. 4.— Redshift slices centered (a) in front of, (b) on, and (c) behind the DLA for the APM 08279+5255 field. The value $z_{\text{abs}}=2.974$ is indicated by the vertical dashed line. Each dot represent a galaxy that was detected in the four UVB& I bands. The filled squares indicate objects that are not detected in the U band. The left column shows the probability distribution as a function of photometric redshift. The continuous line shows the smoothed distribution (arbitrarily scaled to the peak). The right column shows the probability to be in that particular slice as a function of the ‘goodness’ of the photometric redshift $P_{\Delta z}$ defined in Eq. 4. The dotted line shows the minimum threshold (50%) used in selecting LBG candidates in each of the slices.

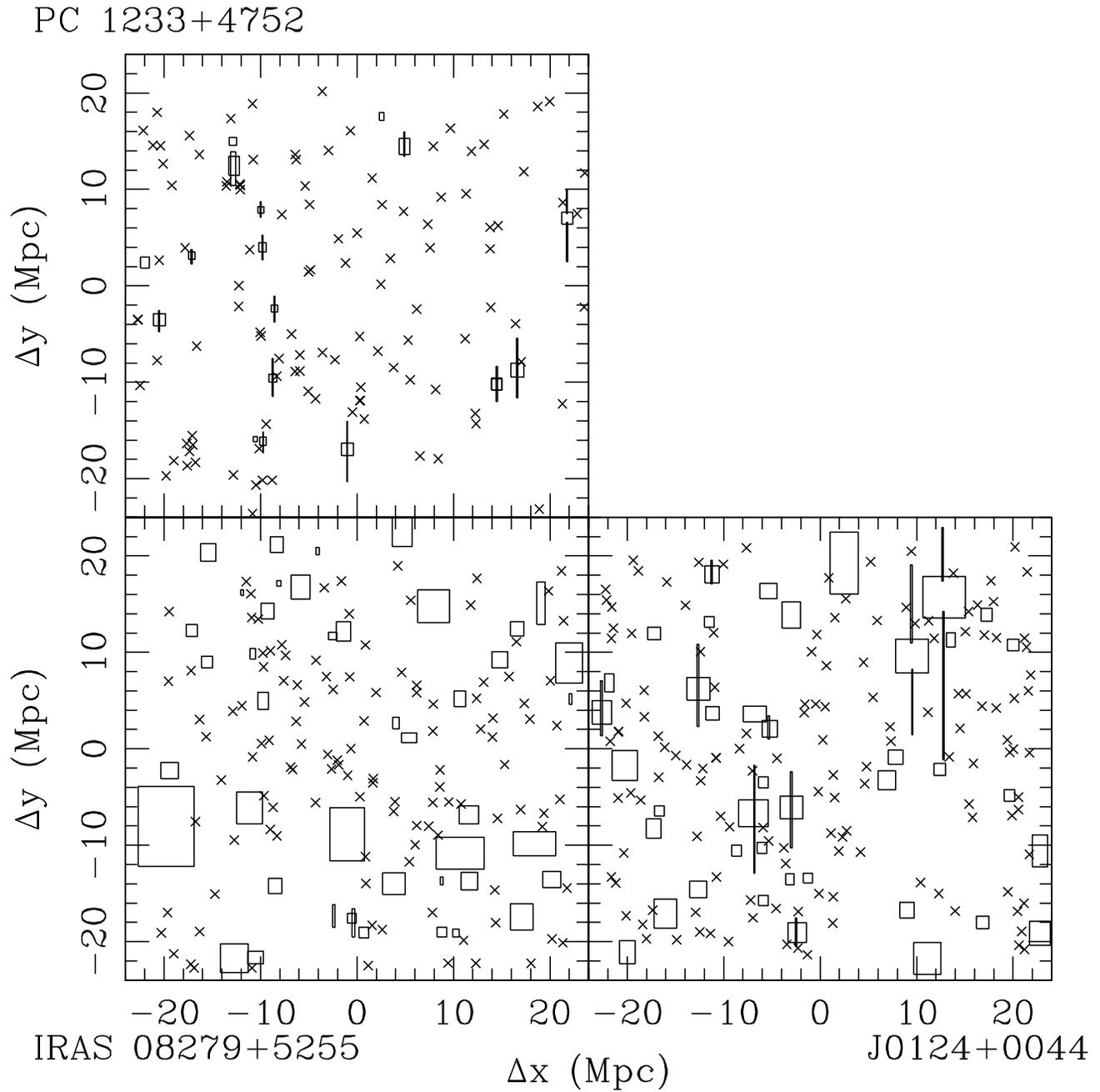


FIG. 5.— For our three fields, the $x y$ position of our LBG candidates relative to the QSO location. North is left, East is down.

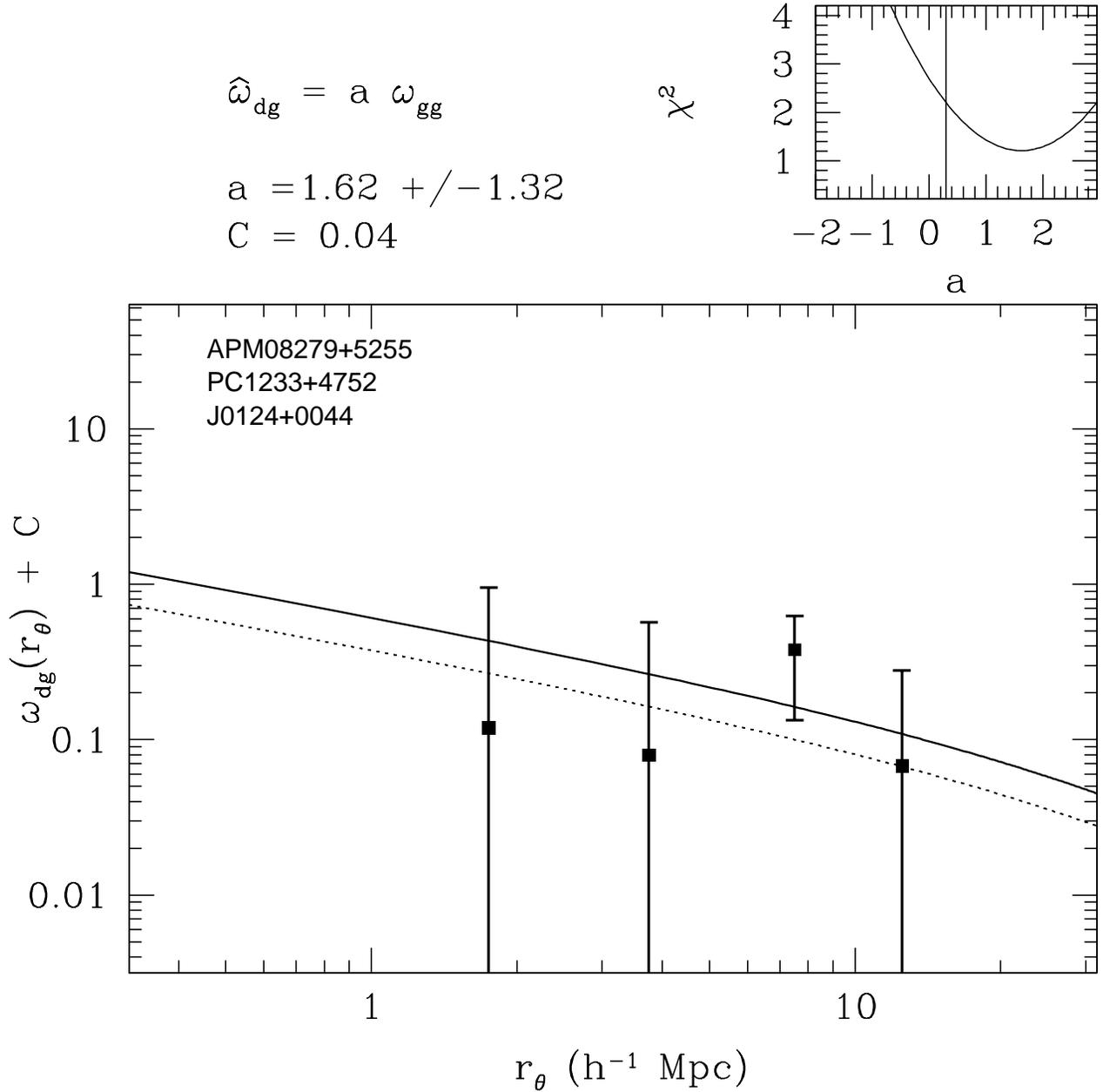


FIG. 6.— The cross-correlation w_{dg} between DLAs and Lyman break galaxies in a redshift slice of width ($W_z = 0.15$) that contains the DLAs. The filled squares show the cross-correlation for the combined fields. The dotted line is the LBG auto-correlation w_{gg} (from Adelberger et al., 2003, using Eq. 8 to account for the volume of our redshift slice). The continuous line is a fit to the amplitude of the cross-correlation using $\hat{w}_{dg} = a \times w_{gg}$, i.e. we assume that both w_{gg} and w_{dg} have the same slope. The small panel shows the χ^2 distribution as a function of the amplitude a and the 1σ range.

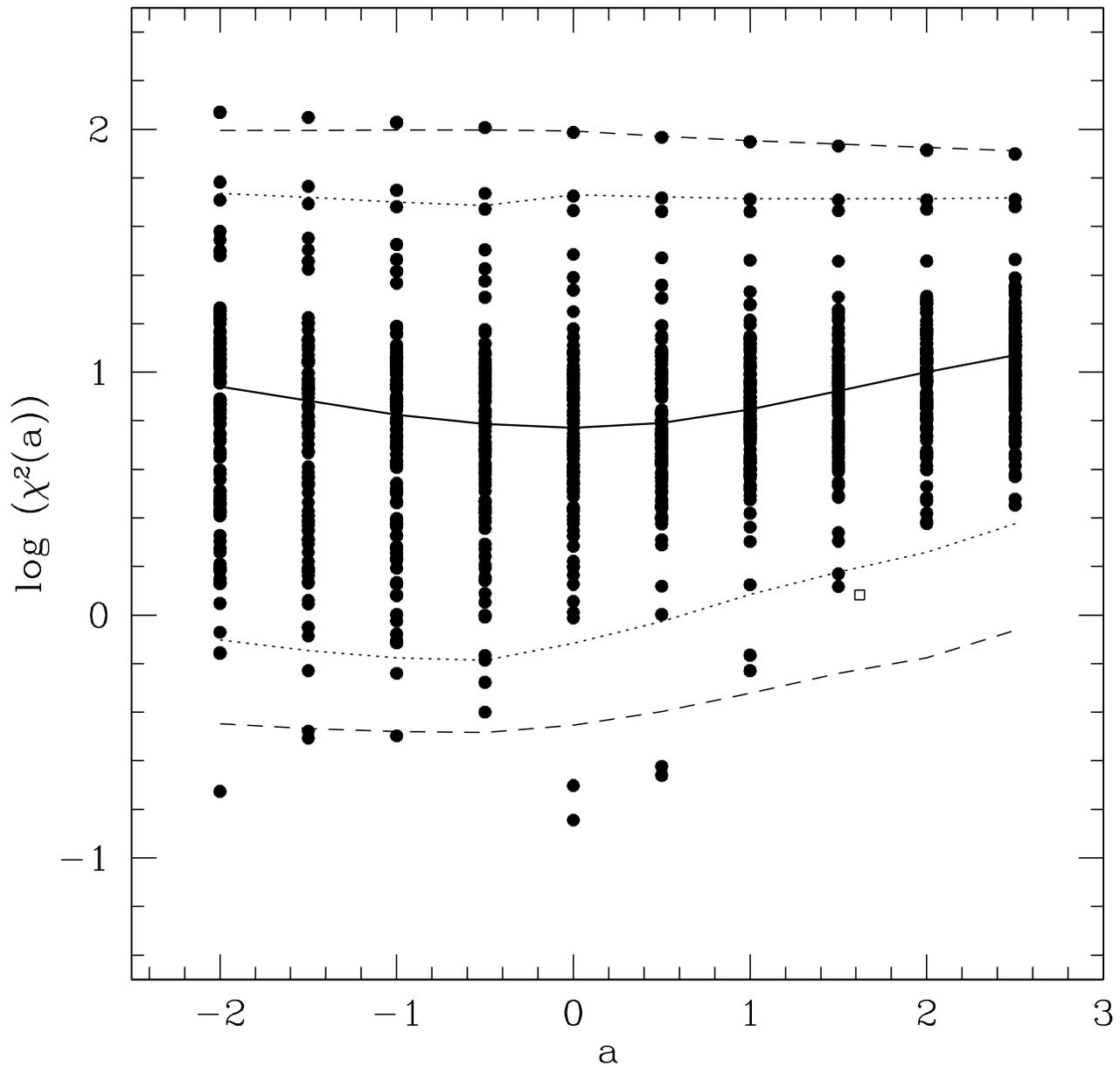


FIG. 7.— The points show the logarithm of $\chi^2(a)$ as a function of the amplitude a for 100 random lines of sight (excluding the central $5h^{-1}$ Mpc) in the redshift slice centered on the DLAs. The continuous line shows the median of the distributions. The dotted line and the dashed line are the 95% and 99% confidence levels, respectively. The open square shows the location of the fitted amplitude shown in Fig. 6, which shows that the signal measured in Fig. 6 is not drawn from a random distribution of lines of sight at the $> 95\%$ confidence level.