We investigate the clustering properties of 2600 Lyman break galaxies (LBGs) at $z = 3.5-5.2$ in two large blank fields, the Subaru Deep Field and the Subaru/XMM-Newton Deep Field (600 arcmin$^2$ each). The angular correlation functions of these LBGs show a clear clustering at both $z \approx 4$ and 5. The correlation lengths are $r_0 = 4.1^{\pm0.2}_{-0.3}$ and $5.9^{+1.3}_{-1.1}h_{100}^{-1}$ Mpc ($r_0 = 5.1^{+1.0}_{-3.1}$ and $5.9^{+1.3}_{-1.7}h_{100}^{-1}$ Mpc) for all the detected LBGs (for $L \geq L^*$ LBGs) at $z \approx 4$ and 5, respectively. These correlation lengths correspond to galaxy–dark matter biases of $b_0 = 2.9^{+0.1}_{-0.1}$ and $4.6^{+0.9}_{-1.2}$ ($b_0 = 3.2^{+0.6}_{-0.7}$ and $4.6^{+0.9}_{-1.2}$) for all the detected LBGs (for $L \geq L^*$ LBGs) at $z \approx 4$ and 5, respectively. These results, combined with estimates for $z \approx 3$ LBGs in the literature, show that the correlation length of $L \geq L^*$ LBGs is almost constant, $5 h_{100}^{-1}$ Mpc, over $z \approx 3-5$, while the bias monotonically increases with redshift at $z \approx 3$. We also find that for LBGs at $z \approx 4$ the clustering amplitude increases with UV-continuum luminosity and with the amount of dust extinction. We estimate the mass of dark halos hosting various kinds of high-$z$ galaxies, including LBGs, with the analytic model given by Sheth & Tormen. We find that the typical mass of dark halos hosting $L \geq L^*$ LBGs is about $10^{12} h^{-1}_7 M_\odot$, over $z \approx 3-5$, which is comparable to that of the Milky Way. A single dark halo with $\sim 10^{12} h^{-1}_7 M_\odot$ is found to host 0.1–0.3 LBGs on average but about four $K$-band–selected galaxies.

Subject headings: cosmology: observations — early universe — galaxies: evolution — galaxies: high-redshift — large-scale structure of universe

1. INTRODUCTION

The formation history of galaxies is basically understood through two fundamental evolutionary processes, i.e., the production of stars and the accumulation of dark matter. Electromagnetic radiation from galaxies gives information about the status of baryonic matter, i.e., stars, gas, and dust. Although dark matter should play a key role in galaxy formation, it cannot be directly detected using electromagnetic waves. The clustering properties of galaxies are closely related to the distribution and amount of the underlying dark matter (e.g., Peacock et al. 2001; Percival et al. 2001, 2002; Verde et al. 2002; Scranton et al. 2002; Connolly et al. 2002; Dodds et al. 2002; Tegmark et al. 2002; Zehavi et al. 2002; Szalay et al. 2003). An analysis of the power spectrum from 2dFGRS data shows that the distribution of galaxies agrees with that of dark matter (Verde et al. 2002). In other words, optically selected galaxies trace the underlying mass distribution at $z = 0$.

The clustering properties of galaxies are now investigated at higher redshifts up to $z \approx 4$. Daddi et al. (2000) have reported a strong angular correlation for extremely red objects (EROs) at $z \approx 1$ and estimated the correlation length of the spatial correlation function to be $r_0 = 12 \pm 3 h^{-1}_{100}$ Mpc, which is comparable to that of present-day ellipticals (Daddi et al. 2001). McCarthy et al. (2001) found similarly large, but slightly smaller, correlation lengths ($r_0 \approx 9-10 h^{-1}_{100}$ Mpc) for their red galaxies at $z \approx 1$. Miyazaki et al. (2003) found strong clustering for both old galaxies and dusty star-forming galaxies beyond $z \approx 1$. Giavalisco et al. (1998) have studied the correlation functions of Lyman break galaxies (LBGs) on the basis of a large sample for bright ($R < 25.5$) LBGs lying around $z = 3$ selected by $U_{10}, GR$ colors. They found that the spatial distribution of $z = 3$ LBGs is strongly biased relative to the dark matter distribution predicted by CDM models, with a linear bias of 4.5 for an Einstein–de Sitter cosmology (see also Adelberger et al. 1998). In a subsequent paper, Giavalisco &
Dickinson (2001) found the clustering amplitude of $z \sim 3$ LBGs to depend on their rest-frame UV luminosity, with fainter galaxies less strongly clustered, which is a property similar to that found in the present-day universe. Arnouts et al. (1999, 2002) have measured the correlation length of faint galaxies in the Hubble Deep Field–North (HDF-N) and –South (HDF-S) over the redshift range $0 < z < 4$, although errors in their measurements are large because of small number statistics. Recent observational results show that LBGs at $z \sim 4$ and Ly$\alpha$ emitters (LAEs) at $z = 4.9$ are strongly biased against the underlying dark matter and that the clustering amplitudes of LAEs are segregated with respect to Ly$\alpha$ luminosity (Ouchi et al. 2001; Ouchi et al. 2003, hereafter Paper II).

Narrowband searches for LAEs on targeted fields have found strong clustering of LAEs around quasars (Campos et al. 1999; Møller & Fynbo 2001; Stiavelli et al. 2001; Fynbo et al. 2003), radio galaxies (Venemans et al. 2002), high-redshift clusters (Palunas et al. 2000), and overdensities in the LBG distribution (Steidel et al. 2000). Recent deep and wide-field LAE searches with the Subaru Suprime-Cam have revealed a large-scale distribution of LAEs (Paper II), and Shimazaki et al. (2003, hereafter Paper IV) have found a filamentary large-scale structure made of LAEs at $z = 4.9$ whose width and length are 20 and more than $50 \ h^{-1}_{70}$ Mpc, respectively. The size of this high-$z$ large-scale structure is comparable to those of present-day large-scale structures found by Geller & Huchra (1989). The existence of such a high-$z$ large-scale structure also indicates that galaxies at $z \sim 5$ would be highly biased against the underlying dark matter. However, LAE searches by narrowband imaging observe a thin slice of the high-$z$ universe, resulting in a surveyed volume of as small as about an order of $10^4 \ h^{-3}_{70}$ Mpc$^3$. Such surveys may suffer from a strong field variance. On the other hand, survey volumes for LBGs are generally about 20–50 times larger than those for LAEs. Thus, measuring the clustering properties of LBGs is highly desirable for quantifying the average distribution of high-$z$ galaxies.

Motivated by this, we carried out deep and wide-field imaging for two blank fields, the Subaru Deep Field (SDF: R.A. = $13^h24^m21^s.4$, decl. = $+27^\circ29'23''$ [J2000.0]; Maihara et al. 2001, hereafter Paper I) and the Subaru/XMM-Newton Deep Field (SXDF: R.A. = $2^h18^m00^s.0$, decl. = $-5^\circ00'00''$ [J2000.0]; K. Sekiguchi et al. 2004, in preparation; see also Ouchi et al. 2001), and made samples of 2600 LBGs at $z = 3.5–5.2$ distributed in these two fields (600 arcmin$^2$ each). Details of the samples are described in Ouchi et al. (2004, hereafter Paper V).

In this paper, we investigate the clustering properties of LBGs at $z = 4$ and 5 using these large LBG samples. We then compare the observational results with the predictions of CDM models as follows: Applying analytic models for the spatial clustering of dark matter to the observational results, we estimate the mass of dark halos that host galaxies from the bias of galaxy–dark matter distribution (Bullock et al. 2002; Moustakas & Somerville 2002). Since the dark-halo mass is a fundamental property of galaxies, we next compare various high-$z$ galaxies detected at various wavelengths in terms of halo mass. In reality, a large fraction of high-$z$ galaxies detected to date are either LBGs, LAEs, EROs, or SCUBA sources.

These high-$z$ galaxy samples are biased toward UV-bright star-forming galaxies (LBGs), strong LAEs, old passive galaxies (EROs), and dusty starburst galaxies (SCUBA sources). It is essential to understand the nature of the high-$z$ galaxy population as a whole by combining these biased samples in order to propose a scenario for the evolution of galaxies at high redshifts. Finally, we examine the relation between high-$z$ galaxies and present-day galaxies on the basis of the dark-halo mass of descendants predicted by the CDM model (Moustakas & Somerville 2002).

The outline of this paper is as follows: In § 2 we describe photometric samples of LBGs at $z = 3.5–5.2$ from the data in the SDF and the SXDF, which are described in Paper V. These are the largest samples of LBGs at $z \geq 4$ in contiguous areas obtained to date and thus enable detailed studies of the clustering properties of galaxies at the highest redshifts so far. Section 3 presents observational results. We investigate the spatial distributions of LBGs to obtain the observational results. We derive the angular correlation functions of the LBGs and calculate the correlation lengths of spatial clustering. We estimate the galaxy–dark matter biases of LBGs at $z \geq 4$ and 5 and compare them with those of low-$z$ galaxies and LAEs at $z = 3$. We examine the dependence of clustering amplitude on luminosity and color for LBGs and LAEs. Section 4 presents theoretical implications. We compare our observational results for the correlation function and the luminosity function (LF) obtained in Paper V with an analytic CDM model to investigate the properties of dark halos hosting high-$z$ galaxies and the relation between galaxies (luminous matter) and dark halos. In § 4 we also infer the masses of the descendants of these high-$z$ galaxies. In § 5 we propose a unified view for a variety of high-$z$ galaxies and discuss the formation history of galaxies from $z = 5$ to 0, based on the results obtained in this paper and Paper V. Section 6 presents a summary of this paper.

Throughout this paper, magnitudes are in the AB system (Oke 1974; Fukugita et al. 1995). The values for the cosmological parameters adopted in this paper are ($\Omega_m$, $\Omega_\Lambda$, $n$, $\sigma_8$) = (0.3, 0.7, 1.0, 0.9). These values are the same as those obtained from the latest cosmic microwave background observations (Spergel et al. 2003). We adopt $n = 1$ and $\sigma_8 = 0.9$ because these values were used in many of the theoretical and observational studies to date. We express physical quantities using $h_{70}$, where $h_{70}$ is the Hubble constant in units of $70$ km s$^{-1}$ Mpc$^{-1}$. The exception is the correlation length, $r_0$. We express $r_0$ using $h_{100}$, where $h_{100}$ is the Hubble constant in units of $100$ km s$^{-1}$ Mpc$^{-1}$, since most of the previous studies use $h_{100}$ to express $r_0$.

2. OBSERVATIONS AND GALAXY SAMPLES

We use the three LBG samples presented in Paper V. Details of the samples are given there, and we give a brief summary of the data and the samples in the following. In §§ 3 and 4 we describe the clustering properties of a sample of LAEs at $z = 4.86 \pm 0.03$ found in the SDF to compare them with those of LBGs. Details of the LAE sample and its clustering properties (e.g., the distribution and angular correlation function) are given in Paper II.

2.1. Observations

During the commissioning runs of Suprime-Cam from 2000 November–2001 November, we carried out multiband,
TABLE 1
PHOTOMETRIC SAMPLES OF GALAXIES

| Field Name | Sample Name | Detection Band | Number | Magnitude Limit $^a$ |
|------------|-------------|----------------|--------|----------------------|
| SDF        | BRi-LBG     | $i'$           | 1438   | $i' < 26.3$          |
| SDF        | Viz-LBG     | $z'$           | 246    | $z' < 25.8$          |
| SDF        | Riz-LBG     | $z'$           | 68     | $z' < 25.8$          |
| SXDF       | BRi-LBG     | $i'$           | 732    | $i' < 25.8$          |
| SXDF       | Viz-LBG     | $z'$           | 34     | $z' < 25.3$          |
| SXDF       | Riz-LBG     | $z'$           | 38     | $z' < 25.3$          |

$^a$ Total magnitudes.

We apply these selection criteria to our photometric catalogs and find 1438 (732), 246 (34), and 68 (38) objects for BRi-LBGs, Viz-LBGs, and Riz-LBGs in the SDF (SXDF), respectively. The $V$-band image of the SXDF is not deep enough to produce a Viz-LBG sample with a small number of contaminants with equation (2). In order to avoid a high contamination rate, we adopt

$$V - i' > 1.2, \quad i' - z' < 0.7, \quad V - i' > 1.8(i' - z') + 2.3$$

(4)

as the Viz-LBG selection criteria for the SXDF. Table 1 summarizes the LBGs found in our data.

We estimate the redshift distribution, completeness, and contamination of the LBG samples using the probability maps obtained from the simulations. The redshift ranges are found to be $z = 4.0 \pm 0.5$ for BRi-LBGs, $z = 4.7 \pm 0.5$ for Viz-LBGs, and $z = 4.9 \pm 0.3$ for Riz-LBGs. Figure 1 shows the number-weighted redshift distributions. The ratio of contaminants to sample objects in number is 1% for BRi-LBGs, 26% for Viz-LBGs, and 40% for Riz-LBGs in the SDF. The completeness of each sample is 40%–60% in the most efficient redshift range. The small ratio of contaminants in BRi-LBGs is due to the fact that the BRi-LBG sample includes many faint LBGs beyond $M^*$, which are much more numerous than foreground contaminants.

2.2. Definition of BRi-, Viz-, and Riz-LBGs

We make three photometric samples of LBGs by the following two color selections: The first sample is for BRi-LBGs at $z \sim 4$. Their Lyman break enters into the $B$ band, and thus they are identified by red $B - R$ and blue $R - i'$ colors. Similarly, the second is for Viz-LBGs at $z \sim 5$ selected by $V - i'$ and $i' - z'$, and the third is for Riz-LBGs at $z \sim 5$ selected by $R - i'$ and $i' - z'$. We define the selection criteria for these LBGs so that the completeness is sufficiently high and the contamination is negligibly small. In order to determine the selection criteria in the two-color diagrams, we use the best-fit spectral energy distributions of the galaxies given in the Furusawa et al. (2000) HDF-N photometric-redshift (photo-z) catalog. We generate artificial galaxies that mimic the colors of the HDF-N galaxies and distribute them randomly on our original images after adding Poisson noise. Then, we detect these simulated objects and measure their brightness in the same manner as for our original images. We iterate this process 100 times and generate probability maps of low-$z$ interlopers and LBGs in the two-color diagrams. Based on the probability maps, we determine the selection criteria for LBGs that give a small number of contaminants and keep a sufficiently high completeness. The selection criteria adopted are

$$B - R > 1.2, \quad R - i' < 0.7,$$

$$B - R > 1.6(R - i') + 1.9 \quad \text{for BRi-LBGs},$$

$$V - i' > 1.2, \quad i' - z' < 0.7,$$

$$V - i' > 1.8(i' - z') + 1.7 \quad \text{for Viz-LBGs},$$

$$R - i' > 1.2, \quad i' - z' < 0.7,$$

$$R - i' > 1.0(i' - z') + 1.0 \quad \text{for Riz-LBGs}.$$
The surface density of LAEs has a large gradient distribution obtained by Papers V and II. Objects identified by two selection criteria but also ensures the incidence not only explains the relatively small numbers of the actual observed numbers within Poisson error. This co-
given in the fourth column of Table 2 and are consistent with PaperI If for LAEs. See PaperV for details. The estimate

distribution derived in Paper V for LBGs and those derived in Paper V. The solid and dashed lines indicate the number-weighted completeness of the LBG samples in the SDF and the SXDF, respectively. Since

the tighter selection criteria (eq. [4]) are applied to Viz-LBGs of the SXDF, their selection window is narrower than that of the Viz-LBGs in the SDF.

Viz-LBGs at z = 4.7 ± 0.5, Riz-LBGs at z = 4.9 ± 0.3, and LAEs at z = 4.86 ± 0.03. Table 2 shows the numbers of objects identified by two selection criteria. These numbers appear quite small. We examine whether these numbers are reasonable. The number of objects selected by two criteria is predicted by

\[ N_{\text{EST}} \simeq \left[ N(\text{SEL1}) - N_c(\text{SEL1}) \right] \frac{\Delta z(\text{SEL2})}{\Delta z(\text{SEL1})} \bar{c}(\text{SEL2}), \quad (5) \]

where \( N(\text{SEL1}) \) is the total number of objects selected by the first criterion, \( N_c(\text{SEL1}) \) is the total number of contaminants in the objects selected by the first criterion, \( \Delta z(\text{SEL2}) \) and \( \Delta z(\text{SEL1}) \) are the approximate redshift ranges of the second and the first criteria, and \( \bar{c}(\text{SEL2}) \) is the number-weighted mean completeness of the second criterion. For this calculation, we use the contamination, completeness, and redshift distribution derived in Paper V for LBGs and those derived in Paper II for LAEs. See Paper V for details. The estimated numbers of objects identified by two selection criteria are given in the fourth column of Table 2 and are consistent with the actual observed numbers within Poisson error. This coincidence not only explains the relatively small numbers of objects identified by two selection criteria but also ensures the accuracy of the contamination, completeness, and redshift distribution obtained by Papers V and II.

3.2. Tests for Detection Inhomogeneity

Figures 2–5 show somewhat inhomogeneous distributions of LBGs. The surface density of LAEs has a large gradient over the whole image. Prior to investigating their clustering properties, we examine whether or not these inhomogeneities come from spatial inhomogeneities of the detection efficiency and/or photometric accuracy in the images. We find that the photometric zero points are accurate within 0.1 mag over the whole field of view (FOV) for any bandpass, since point-spread function–like objects are found to make a single sharp stellar locus in any two-color plane, as shown in Paper V.

Then we examine whether there are spatial differences in the source detection efficiency by (1) calculating the number densities of all detected objects in small (10′′ × 10′′) areas covering the survey region, (2) measuring the limiting magnitudes in 2700 small (40′′ × 40′′) areas for each of the B, V, R, i′, z′, and NB711 images, and (3) estimating the detection completeness of LBGs and LAEs from Monte Carlo simulations in the same manner as in Paper II. We find no clear sign of inhomogeneity in any of points 1−3 for the SDF data. On the other hand, we find a systematic difference in the detection limits for the SXDF data at a 0.2 mag level on arcminute scales. This is probably because the majority of the SXDF data were taken in 2000, when the CCDs installed were a mixture of products of two companies and so had a large variety in quantum efficiency. Figures 2–5 show, however, that there is no correlation between the source distribution and the CCD positions. This demonstrates that the inhomogeneities seen in Figures 2–5 are real, although the distributions of LBGs in the SXDF on arcminute scales could be
contaminated by the somewhat large inhomogeneities in the limiting magnitude.

3.3. Angular Correlation Function

In order to quantitatively measure the inhomogeneity of the spatial distribution, we derive the angular two-point correlation function $\omega(\theta)$. According to Landy & Szalay (1993), the angular two-point correlation function is calculated by

$$\omega_{\text{obs}}(\theta) = \frac{DD(\theta) - 2DR(\theta) + RR(\theta)}{RR(\theta)},$$

where $DD(\theta)$, $DR(\theta)$, and $RR(\theta)$ are the numbers of galaxy-galaxy, galaxy-random, and random-random pairs normalized by the total number of pairs in each of the three samples. We create a random sample composed of 100,000 sources with the same geometrical constraints as the data sample. The formal error in $\omega(\theta)$ is described by

$$\sigma_{\omega} = \sqrt{1 + \omega_{\text{obs}}(\theta)} / DD(\theta).$$

(Hewett 1982). Its 2 $\sigma$ error is comparable to the error obtained with a bootstrap resampling of the data (Baugh et al. 1996). The real correlation function $\omega(\theta)$ is offset from the observed function by an integral constant (IC; Groth & Peebles 1977) as

$$\omega(\theta) = \omega_{\text{obs}}(\theta) + \text{IC}.$$  

We apply the correction for the IC.\footnote{If we assume $\beta = 0.8$, IC/$A^2(\theta)$ is estimated to be $\geq 0.006$ for our LBG samples, where $A^2(\theta)$ and $\beta$ are the correlation amplitude and the power-law index shown in eq. (9).}

Figure 6 shows the angular correlation functions $\omega(\theta)$ of $BRi$-LBGs (top), $Viz$-LBGs (middle), and $Riz$-LBGs (bottom). We measure the angular correlation function at less than $\sim 15'$ scales, since uncertainties in the measurements increase largely over $\sim 15'$ scales, which are comparable to half of the image size. In Figure 6 the filled circles are for the SDF, and the open circles in the top panel are for the SXDF. We find significant clustering signals for $BRi$-LBGs in both the SDF and the SXDF. The data points of the SDF and SXDF samples agree...
within the 2 $\sigma$ level except for the points at $\theta \simeq 30''$ and $\simeq 140''$, where the measurements of the SXDF are significantly lower than those of the SDF. The reason for these discrepancies is not clear. However, it could be related to the poorer quality of the SXDF data compared to the SDF data. In any case, the SDF sample is deeper than the SXDF sample and does not suffer from any detectable (artificial) inhomogeneity in the data. Hence, we believe that the angular correlation measured for the SDF sample is more reliable. We also find clear clustering signals for $Riz$-LBGs in the SDF. There is a marginal signal for $Viz$-LBGs in the SDF, and no signal is detected for $Viz$-LBGs and $Riz$-LBGs in the SXDF. This is mainly because the numbers of objects in these samples are considerably smaller than those in the other samples.

We fit a single power law, 

$$\omega(\theta) = A_{\text{raw}}^{\text{raw}} \theta^{-\beta},$$

(9)

to the data points. For $BRi$-LBGs, a fit is made to a combination of the data points of the SDF and the SXDF. (A fit to the SDF data alone gives almost the same results, since the errors in the data points of the SDF sample are much smaller.) Errors in $A_{\text{raw}}$ and $\beta$ are found to be very large except for $BRi$-LBGs: $\beta = 0.90^{+0.11}_{-0.07}$ and $A_{\text{raw}} = 2.3^{+0.9}_{-1.4}$ (Table 3, row [1]). This is probably due to the large sample size of $BRi$-LBGs. The value of $\beta$ for $BRi$-LBGs is close to the "fiducial" value, 0.8. Thus, we use $\beta = 0.8$ in the rest of this paper. Adopting $\beta = 0.91$ changes the results very little. Rows (2)–(14) of Table 2 present the best-fit values of $A_{\text{raw}}$ with $\beta = 0.8$. The correlation amplitudes $A_{\text{raw}}$ of $BRi$-LBGs and $Viz$-LBGs are as high as the $\sim 10$ and $\sim 2.5$ $\sigma$ levels, respectively. Hence, their clustering signals are significant. On the other hand, the $A_{\text{raw}}$ value of $Riz$-LBGs is only at the less than 1 $\sigma$ significance level. Thus, in the following discussion, we adopt the $A_{\text{raw}}$ value of SDF $Viz$-LBGs for the clustering amplitude of $z = 5$ LBGs.

Foreground contamination in a galaxy sample dilutes the apparent clustering amplitude of the galaxies. When the fraction of contaminants is $f_c$, the apparent $A_{\text{raw}}$ value can be reduced by a factor of up to $(1 - f_c)^2$. The true correlation amplitude, $A$, is given by 

$$A = \frac{A_{\text{raw}}}{(1 - f_c)^2}.$$

(10)

This is the maximum reduction of the correlation amplitude that occurs when the contaminants are not at all clustered. In reality, the contaminants in our sample will be the sum of foreground galaxies at various redshifts and thus will be clustered very weakly, if at all, on the sky. Thus, we use equation (10) to compute $A_f$ for our LBG samples. We use the $f_c$ obtained by the simulations described in Paper V for the LBGs. Table 3 also gives the $A_f$ values.

The effect of field-to-field variations in our samples is probably modest for LBGs, since we find that angular correlation functions of $BRi$-LBGs in the SDF and the SXDF agree moderately well within error bars (Fig. 6). This is because our LBG samples probe large comoving volumes: $1.6 \times 10^6 h^{-3} \text{ Mpc}^3$ for $BRi$-LBGs and $1.7 \times 10^6 h^{-3} \text{ Mpc}^3$ for $Viz$-LBGs. On the other hand, the surveyed volume of LAEs at $z = 4.9$ shown in Paper II is only $9.0 \times 10^5 h^{-3} \text{ Mpc}^3$. We may have to consider cosmic variance for the LAE sample.

Since we have a large number of LBGs at $z = 4$ in the SDF, we make subsamples that are divided by magnitude or color. We use two observational quantities, the $i'$ magnitude and $i' - z'$ color, and make three types of subsamples: a subsample composed of objects with $M \leq M^*$, subsamples divided by $M$, and subsamples divided by $E(B - V)$. Then we calculate the angular correlation functions of these subsamples in the same manner as shown above for the entire LBG sample. We describe the details of the analysis in §3.5.3 and 3.6 and summarize the selection criteria and the results in Table 3. We do not apply this analysis to the SXDF $BRi$-LBG sample, since its limiting magnitude is brighter than that of the SDF sample (and so the size of the SXDF sample is about half that of the SDF sample), and the SXDF sample could suffer from the inhomogeneity of the limiting magnitude over the image.

### 3.4. Correlation Length

Since the angular correlation function shows the clustering properties of galaxies projected on the sky, $\omega(\theta)$ reflects a
combination of the redshift distribution of the selected galaxies and the intrinsic clustering in three-dimensional space, i.e., the spatial correlation function \( \xi_R \). The spatial correlation function of galaxies is usually expressed by a power law as

\[
\xi_R = (r/r_0)^{-\gamma},
\]

where \( r \) is the spatial separation between two objects, \( r_0 \) is the correlation length, and \( \gamma \) is the slope of the power law. The correlation length \( r_0 \) is related to the correlation amplitude \( A_\xi \) with the integral equation called the Limber equation (Peebles 1993),

\[
A_\xi = C \int_0^\infty F(z)D_0^{-1}(z)(z)^2 g(z)\,dz \left[ \int_0^\infty N(z)\,dz \right]^{-2},
\]

where \( F(z) \) describes the redshift dependence of \( \xi(r) \), \( D_0(z) \) is the angular diameter distance, \( N(z) \) is the redshift distribution of objects,

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

and \( C \) is a numerical constant, \( C = \sqrt{\pi} \Gamma[(\gamma - 1)/2]/\Gamma(\gamma/2) \).

The slope \( \beta \) of the angular correlation function is related to \( \gamma \) by

\[
\gamma = \beta + 1.
\]

We need the redshift distributions of our LBGs to derive \( r_0 \). We adopt the ones obtained in § 2.

The correlation lengths thus obtained are summarized in Table 4, where we show both \( r_0 \) and \( r_{0\text{raw}} \), which are the correlation lengths obtained from \( A_\xi \) and \( A_{\xi,\text{raw}} \). In the following discussion, we regard \( r_0 \) as the best estimate value. We find \( r_0 = 4.1^{+0.2}_{-0.2} \) and \( 5.9^{+1.3}_{-0.7} h_{100} \) Mpc for all LBGs at \( z = 4 \) and \( 5 \), respectively. Ouchi et al. (2001) calculate the correlation length of LBGs at \( z = 4 \) in the same manner as above but using a sample constructed from different selection criteria. They report that LBGs at \( z = 4 \) with \( i' < 26 \) have \( r_0 = 3.3^{+0.6}_{-0.7} h_{100} \) Mpc (\( r_{0\text{raw}} = 2.7^{+0.6}_{-0.5} h_{100} \) Mpc). This value is consistent with that obtained from our LBG sample. The correlation lengths we obtain \( (r_0 = 3.8 h_{100} \text{ Mpc}) \) correspond to projected angular scales of \( 100' \text{ to } 300' \) at \( z = 4-5 \). These scales fall in the range in which the power-law fit is excellent (Fig. 6).

3.5. Evolution of Correlation Length and Galaxy Distribution Bias

We discuss the redshift evolution of the correlation length. Since the clustering amplitude (i.e., correlation length) depends on the brightness of the galaxies (see § 3.6.1), we compare the correlation lengths of LBGs whose luminosities are \( L \geq L^* \). We calculate the correlation length of LBGs with \( L \geq L^* \) for \( z = 4 \) and \( 5 \). We make a subsample composed of 357 LBGs at \( z = 4 \) with \( i' < 25.3 \) corresponding to \( M < -20.8 \) (\( \sim M^* \)) and calculate the correlation length to be \( r_0 = 5.1^{+1.0}_{-1.1} h_{100} \) Mpc (row [2] of Table 5). Since the limiting magnitude of \( z = 5 \) LBGs is \( z' = 25.8 \), which corresponds to \( M = -20.5 \) (\( \sim M^* \)), we regard the results of all \( z = 5 \) LBGs (row [2] of Table 4) as those of \( L \geq L^* \) LBGs (row [3] of Table 5). Thus, the correlation length of \( L \geq L^* \) LBGs at \( z = 5 \) is \( r_0 = 5.9^{+1.3}_{-1.7} h_{100} \) Mpc. We also show in Table 5 the correlation
lengths of LBGs with $L \sim L^*$ at $z = 3$ given by Giavalisco & Dickinson (2001) and of all LAEs at $z = 4.9$ obtained by Paper II.

Figure 7 (top) plots $r_g$ as a function of redshift, together with the $r_g$ of the underlying dark matter calculated by the nonlinear model of Peacock & Dodds (1996). The correlation lengths of LBGs with $L \sim L^*$ is almost constant around $r_0 \sim 5 h^{-1}_{100}$ Mpc at $z = 3-5$, although it might increase slightly with redshift. It is also found that the correlation amplitudes of LBGs (and LAEs) are much higher than that of the underlying dark matter. We define the galaxy–dark matter clustering bias at the large scale ($\sim 8 h^{-1}_{100}$ Mpc) as

$$b_g = \sqrt{\xi_g(r = 8 h^{-1}_{100} \text{ Mpc})/\xi_{DM}(r = 8 h^{-1}_{100} \text{ Mpc})}$$  \hspace{1cm} (14)$$

where $\xi_{DM}$ is the two-point correlation function of the underlying dark matter.

We also calculate the $r_0$ of the dark matter using linear theory (e.g., Peacock & Dodds 1994), which is plotted in Figure 7 by a dashed line, and find a negligible difference from the calculation based on the nonlinear model. This is because the effect of nonlinearity is very small on scales of $r \geq r_0$. Jenkins et al. (1998) show that linear theory traces the results of $N$-body simulations above $r_0$ at $z = 0$, where the nonlinear effect is larger than at high redshifts.

Using linear theory, we calculate $b_g$ by equation (14) from $r_0$ (eq. [11]), $\gamma$ (eq. [13]), and $\xi_{DM}$. We explicitly present the relation between $r_0$ and $b_g$ given in equation (14):

$$b_g = \left(\frac{8 h^{-1}_{100} \text{ Mpc}}{r_0}\right)^{\gamma}/\xi_{DM}(r = 8 h^{-1}_{100} \text{ Mpc})$$  \hspace{1cm} (15)$$

We show values of $b_g$ in Table 5 for LBGs with $L \sim L^*$ and all LAEs (and Table 4 for all samples), and Figure 7 (bottom) plots $b_g$ against redshift for LBGs with $L \sim L^*$ and all LAEs.

### Table 5

| Sample Name | $(\bar{z})$ | Criteria | $r_0^b$ | $b_g^b$ |
|-------------|-------------|-----------|---------|---------|
| U$_b$GR-LBG (SPEC) | 3.0 | $M_{1700} < -20.5 \sim M^*$ ($R < 25.0$) | $5.0^{+0.7}_{-0.7}$ | $2.7^{+0.4}_{-0.4}$ |
| BRi-LBG | 4.0 | $M_{1700} < -20.8 \sim M^*$ ($i' < 25.3$) | $5.1^{+1.0}_{-0.7}$ | $3.5^{+0.6}_{-0.7}$ |
| Viz-LBG | 4.7 | $M_{1700} < -20.5 \sim M^*$ ($z' < 25.8$) | $5.9^{+1.0}_{-1.7}$ | $4.6^{+0.7}_{-1.2}$ |
| LAE | 4.9 | $M_{1700} < -19.0 (z' < 27.3)$ | $6.2^{+0.5}_{-0.5}$ | $4.9^{+0.4}_{-0.4}$ |

**Notes.**—Values for BRi-LBGs and Viz-LBGs are the same as those presented in Table 4.

* These criteria correspond to $L \sim L^*$ for LBGs at each redshift (see Paper V).

* Values for $r_0$ and $b_g$ for BRi-LBGs, Viz-LBGs, and LAEs are contamination-corrected.

* Values for $r_0$ and $\gamma$ of U-band dropout LBGs are obtained from the SPEC sample of Giavalisco & Dickinson (2001).

* Values for $r_0$ and $b_g$ are obtained from Paper II. These are contamination-corrected values.
These LBGs and LAEs are biased against dark matter by \( b_g \approx 3\)–5, and the bias becomes stronger at higher redshifts. This increase in the LBG bias can be regarded as a piece of evidence supporting the biased galaxy formation scenario (e.g., Baugh et al. 1999).

### 3.6. Segregation of Clustering Amplitudes

#### 3.6.1. Clustering Segregation with UV-Continuum Luminosity

Since we have a large number of LBGs at \( z = 4 \), we examine the luminosity dependence of the clustering amplitude \( r_0 \) and \( b_g \) using four subsamples binned according to the source brightness \( \Delta m = 1.5 \). We calculate the angular correlation functions for the four subsamples and estimate \( r_0 \) and \( b_g \) in the same manner as described in §§ 3.4 and 3.5, but applying to each subsample its own redshift distribution, which is determined by the simulations described in Paper V (Fig. 12 of Paper V). Rows (6)–(9) of Table 3 and rows (4)–(7) of Table 4 give the best-fit parameters of the angular correlation function together with \( r_0 \) and \( b_g \) for these four subsamples. Figure 8 plots \( r_0 \) and \( b_g \) for the four subsamples as a function of magnitude. It is found from Figure 8 that brighter LBGs are clustered more strongly, which is a tendency similar to that found for present-day galaxies (Norberg et al. 2002) and LBGs at \( z = 3 \) (Giavalisco & Dickinson 2001). It is also found that this luminosity dependence of \( r_0 \) is weaker for fainter galaxies. We fit the data points with a linear function of luminosity \( (L) \):

\[
b_g/b_g^* = a + (1 - a)L/L^* ,
\]

where \( a \) is a free parameter, \( b_g \) is the galaxy–dark matter bias, and \( b_g^* \), which is also a free parameter, is the bias of galaxies with luminosity \( L^* \) (Norberg’s law; Norberg et al. 2002). We obtain \( b^* = 2.8 \) and \( a = 0.58 \) with \( L^* \) being fixed at \( M_{*}^{1700} = -21.0 \) (or \( i^* = 25.1 \)), which is obtained in Paper V. The value of \( a \) for \( z = 4 \) LBGs is smaller than that for present-day galaxies, \( a = 0.85 \) (Norberg et al. 2002). It may indicate that the clustering amplitude of \( z = 4 \) LBGs depends more strongly on luminosity and/or that the luminosity dependence of clustering is different between rest-frame UV luminosity (for LBGs) and optical luminosity (for present-day galaxies).

#### 3.6.2. Clustering Segregation with Dust Extinction

Strong clustering has been reported for SCUBA sources \( (r_0 = 12.8 \pm 4.5 \pm 3.0 \ h_{100}^{-1} \) Mpc; Webb et al. 2003). Since SCUBA sources are thought to be dust-rich starburst galaxies, the clustering amplitude of LBGs as a function of dust extinction may give a hint for the relation between SCUBA sources and LBGs. We calculate the clustering amplitude for subsamples of LBGs at \( z = 4 \) binned with \( E(B-V) \). We estimate \( E(B-V) \) of LBGs from their colors with the equations given in Paper V. We make these subsamples from 357 bright LBGs with \( i' < 25.3 \) (\( \approx M^* \)) so that we accurately measure colors, \( i' - z' \), and thus obtain \( E(B-V) \). We make five
subsamples whose \(E(B - V)\) values range from \(-0.2\) to \(0.6\), corresponding to an \(i' - z'\) color from \(-0.18\) to \(0.50\).\(^{16}\) Rows (10)–(14) of Table 3 and rows (8)–(12) of Table 4 present the best-fit parameters of the angular correlation functions, together with \(r_0\) and \(b_g\) for these five subsamples. We find that the most dusty subsample \([0.2 < E(B - V) < 0.6, \text{i.e.,} \ E(B - V) \approx 0.4]\) has the strongest clustering: \(r_0 = 8.2^{\pm 0.2}_{\mp 0.2} h_{100}^{-1}\) Mpc. Although the errors are large, it may suggest that very dusty LBGs have a connection to SCUBA sources.

On the other hand, Ouchi et al. (1999) find that no LBG identified in the HDF-N is detected by deep SCUBA observations (Hughes et al. 1998). There are three possible explanations for this finding. First, dusty LBGs bright enough to be detected with SCUBA may be so rare that the probability of finding such an LBG in a small area such as the HDF-N is very low. Second, the redshift distributions of LBGs and SCUBA sources may overlap little with each other. Indeed, SCUBA sources are generally located at \(z = 1.9 - 2.8\) (Chapman et al. 2003), while LBGs are found at redshifts higher than \(z \approx 2.5\). Third, LBGs may have a closer relation with EROs than with SCUBA sources. EROs are thought to be a mixture of old galaxies and dusty star-forming galaxies. Although we assume that the red \(i' - z'\) color of LBGs is due to dust extinction, the reddest LBG subsample might include a large number of old galaxies that have optical to infrared colors close to those of EROs. In reality, McCarthy et al. (2001) find that the correlation length of EROs at \(z = 1\) is \(9 - 10 h_{100}^{-1}\) Mpc, comparable to that of the reddest LBG subsample.

4. GALAXY AND STRUCTURE FORMATION BASED ON THE HIERARCHICAL CLUSTERING PICTURE

4.1. Joint Analysis of LFs and Clustering Amplitudes based on the CDM Model

We estimate the mass of dark halos that host LBGs and LAEs using the analytic models given by Sheth & Tormen (1999) and Sheth et al. (2001). They are derived from a fit to the results of GIF simulations (GIF/Virgo collaboration; e.g., Kauffmann et al. 1999).

Prior to estimating the mass of dark halos, we present an overview of the model of Sheth & Tormen (1999). According to their model (see also Mo & White 2002), the number density of collapsed dark halos with mass \(M\) at redshift \(z\) is given by

\[
n(M, z) dM = A \left(1 + \frac{1}{\nu'^2}\right) \sqrt{\frac{2}{\pi}} \frac{\rho}{M} d\nu' \exp\left(-\frac{\nu'^2}{2}\right) dM,\]

where \(\nu' = \sqrt{\nu} a, a = 0.707, A \approx 0.322, \text{and} q = 0.3\). Here, \(\nu\) is defined by

\[
\nu \equiv \frac{\delta_c}{D(z) \sigma(M)},
\]

where \(D(z)\) is the growth factor, \(\sigma(M)\) is the rms of the density fluctuations on mass scale \(M\), and \(\delta_c = 1.69\) represents the critical amplitude of the perturbation for collapse. We calculate the growth factor following Carroll et al. (1992), in which

\[
\sigma(M) = \frac{\sqrt{\ln(100) + \left(\frac{1}{2} - 0.18z + 0.59z^2\right)}},
\]

where \(b = 0.5, c = 0.6, \text{and the other parameters (} \nu', \delta_c, \text{and} a \text{) are the same as in equations (17) and (18) (Sheth et al. 2001). Since} \nu' \text{is a function of redshift and mass, the mass of a dark halo is estimated from equations (18) and (19) (Sheth et al. 2001), once the bias of the dark halo and the redshift are given. Since the bias values of LBGs,} b_g, \text{are measured at a large separation of} 8 h^{-1} \text{Mpc, these values reflect the bias of dark halos (which host LBGs) rather than the bias of the galaxy distribution in dark halos, \text{i.e.,} \ b_g \approx b_{DH}. \text{Assuming that galaxies with a given brightness are hosted by dark halos of a specific mass (\text{i.e.,} there is a one-to-one correspondence between the galaxy luminosity and the halo mass), we estimate the average mass of dark halos,} \langle M \rangle, \text{for galaxies of a given luminosity range from the bias values obtained in} \S \ 3.5. \text{We also calculate the number density of galaxies from the LFs given in Paper V. Note that since the LFs are derived by correcting completeness and contamination for the number counts of LBGs, these number densities do not suffer from these systematic observational errors. The ticks in Figure 8 show the halo masses for} z = 4. \text{The ordinate shows either the number density of LBGs in a given dark-halo mass range (circles) or the number density of dark halos (thin solid line), and} \Delta M \text{shown in the ordinate axis means the number density in the mass range of 1 mag, i.e.,} \Delta M = M/\langle 2.5 \log(c) \rangle = 0.92 M. \text{The thin solid line indicates the mass function of all dark halos at} z = 4 \text{calculated using the formula given by Sheth & Tormen (1999). The four circles correspond to the four data points shown in Fig. 8. The thick solid line corresponds to a fit of the scaled Norberg’s law to the four data points of Fig. 8.}
and the number densities for \( z = 4 \) LBGs thus obtained. We then plot in Figure 9 the number density of LBGs against the mass of hosting dark halos for each luminosity bin, together with the mass function of all dark halos existing at \( z = 4 \) calculated by equation (17).

4.2. Hosting Dark Halos and Halo Occupation Number

4.2.1. Dark Halos of LBGs at \( z = 4 \)

Figure 9 shows that dark halos that host LBGs at \( z = 4 \) have \( 10^{11} – 10^{13} h_{70}^{-1} M_\odot \). Each filled circle denotes the mass of the hosting halos for each luminosity bin (see Table 6 for the definition of the luminosity bins); from right to left are halos hosting LBGs with \( i' \simeq 24.0, 24.5, 25.0, \) and 25.5. The third circle from the right is for \( i' \simeq 25.0 \) \((M_{70700} \approx -21)\) LBGs, i.e., for L* LBGs (see Paper V). We find that bright LBGs (\( M \approx M^* – 1 \)) have more massive dark halos (\( 4.6 \times 10^{12} h_{70}^{-1} M_\odot \)) than faint LBGs (\( M \approx M^* + 0.5 \); \( 1.7 \times 10^{11} h_{70}^{-1} M_\odot \)). A qualitatively similar trend has been found for LBGs at \( z = 3 \) (Giavalisco & Dickinson 2001).

The most striking feature in Figure 9 is the discrepancy in number density between all dark halos (thin solid line) and all the data points (circles) except the brightest one. We define the occupation number, \( N_{\text{occ}} \), as the ratio of the number density of galaxies to that of dark halos. Table 6 presents \( N_{\text{occ}} \) for each luminosity bin together with the results for galaxies at \( z = 3 \) and 5, which are explained in the following sections. Rows (10)–(13) of Table 6 (and Fig. 9) show that the occupation number of the brightest LBGs is almost unity, but those of fainter LBGs (including L* LBGs) are 3–10 times smaller. This implies that the majority of low-mass (\( \leq 10^{12} h_{70}^{-1} M_\odot \)) dark halos have no LBGs. Since LBGs are star-forming

| Sample Name | \((z)\) | Selection Criteria | \(\langle M \rangle \) (\(h_{70}^{-1} M_\odot\)) | \(\log (\text{N})\) | \(N_{\text{occ}}\) | \(\langle M \rangle (z = 0) \) (\(h_{70}^{-1} M_\odot\)) | \(N_{\text{min}}^\text{merge} \) (merge) |
|-------------|--------|------------------|----------------------|-----------------|-------------|----------------------|-------------------|
| UaGR-LBG (PHOT) | 0.80 | \( R < 25.5 \) | \(3.0 \times 10^{11} \) | 0.6 | 0.8 | \(1.4 \times 10^{11} \) | 1.5 |
| UaGR-LBG (SPEC) | 0.80 | \( R < 25.0 \) | \(3.0 \times 10^{11} \) | 0.6 | 0.8 | \(1.4 \times 10^{11} \) | 1.5 |
| UBL-LBG (HDF) | 0.80 | \( V_{606} < 27.0 \) | \(3.0 \times 10^{11} \) | 0.6 | 0.8 | \(1.4 \times 10^{11} \) | 1.5 |
| FIREs all | \( K < 24 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| FIREs red | \( K < 24, J – K > 1.7 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| FIREs blue | \( K < 24, J – K < 1.7 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| SCUBA | \(< 3 \) | \(5.0 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( i' \leq 26 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( i' < 25.3 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( 23.3 < i' < 24.8 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( 23.8 < i' < 25.3 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( 24.3 < i' < 25.8 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| BB-LBG | \( 24.8 < i' < 26.3 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |
| VLB-LBG | \( 4.7 < i' < 25.8 \) | \(2.9 \times 10^{10} \) | 0.6 | 0.8 | 1.0 | 1.0 |

Notes.—The quantities \( M(z) (h_{70}^{-1} M_\odot) \) and \( M(z = 0) (h_{70}^{-1} M_\odot) \) are the mass of dark halos hosting high-z galaxies and the typical mass of their descendants at \( z = 0 \), respectively, \( \log (\text{N}) \) is the number density of high-z galaxies in units of \( h^{-1} \) Mpc\(^{-3} \). 

The maximum mass and occupation number of LAEs are \( 3.5\times 10^{12} h_{70}^{-1} M_\odot \) and 35 (see § 4.2.3).

In order to investigate the discrepancy in number density between LBGs and all dark halos (Fig. 9), we carry out the same analyses for various galaxy populations at \( z \approx 3 \), whose number density and correlation length have been reported in the literature. We examine K-band–selected galaxies (FIRES; Daddi et al. 2003), SCUBA sources (Webb et al. 2003), and LBGs (Giavalisco & Dickinson 2001). Here, we note the basic properties of these three populations. FIRES galaxies are detected by very deep K-band imaging (\( K < 24 \)) with the Very Large Telescope ISAAC (Daddi et al. 2003). Although the sample of FIRES galaxies may suffer from field variance due to its very small survey field (\(~ 4 \) arcmin\(^2 \)), Daddi et al. (2003) claim from their simulations that the strong correlation amplitudes they found are hardly reproduced by the field variance alone. FIRES galaxies include UV-faint galaxies, i.e., quiescent in star formation activity, which are not identified by LBG selections (Franx et al. 2003). SCUBA sources are detected by their strong submillimeter fluxes in the Canada-UK Deep Submillimeter Survey with the James Clerk Maxwell Telescope SCUBA, and they are thought to be a mixture of dusty starburst galaxies and active galactic nuclei. The sample of LBGs at \( z = 3 \) used in this analysis is taken from Giavalisco & Dickinson (2001); see also Giavalisco et al. 1998; Adelberger et al. 1998; Arnouts et al. 1999, 2002; Porciani & Giavalisco 2002).

Figure 10 (and rows [1]–[7] of Table 6) shows the results for \( z \approx 3 \) galaxies. The mass of dark halos hosting LBGs ranges over \( 10^{11} – 10^{12} h_{70}^{-1} M_\odot \). The occupation number of LBGs is less than unity (\(~ 0.3 \); see Table 6). These features of LBGs at \( z = 3 \) are similar to those at \( z = 4 \). On the other hand, dark halos of FIRES galaxies have \( \sim 10^{12} h_{70}^{-1} M_\odot \), which is
and large occupation number indicate that dark halos with occupation number of FIRES galaxies is as high as about an order of magnitude larger than for other objects. The two open triangles are for SCUBA sources is about 4, although the uncertainty in it is filled triangle corresponds to all galaxies found in FIRES (Daddi et al. 2003), while the two open triangles are for FIRES galaxies with blue color (left: J − K < 1.7) and red color (right: J − K > 1.7). The star corresponds to SCUBA sources whose correlation length is reported to be \( r_0 = 12.8 \pm 4.5 \, h_{70}^{-1} \) Mpc by Webb et al. (2003) (see Scott et al. 2002 for their number density).

fig. 10.—Same as Fig. 9 but for galaxies at \( z = 3 \). The two filled circles show LBGs at \( z = 3 \) in the SPEC sample (right) and the PHOT sample (left) of Giavalisco & Dickinson (2001). The open circle with an arrow indicates the measurement for the HDF-N sample of Giavalisco & Dickinson (2001). The filled triangle corresponds to all galaxies found in FIRES (Daddi et al. 2003), which is a K-band–limited high-z galaxy survey with \( K < 24 \) (Labbè et al. 2003), while the two open triangles are for FIRES galaxies with blue color (left: J − K < 1.7) and red color (right: J − K > 1.7). The star corresponds to SCUBA sources whose correlation length is reported to be \( r_0 = 12.8 \pm 4.5 \, h_{70}^{-1} \) Mpc by Webb et al. (2003) (see Scott et al. 2002 for their number density).

fig. 11.—Same as Fig. 9 but for LBGs and LAEs at \( z = 5 \). The filled circle and filled square indicate LBGs and LAEs, respectively. The open square is for LAEs, but when a higher dark-halo mass, \( 3.5 \times 10^{12} \, h_{70}^{-1} \) Msun, is adopted. See § 4.2.3 for details.

comparable to the mass of LBGs, but the occupation number of FIRES galaxies is as high as \( \approx 4 \). This comparable mass and large occupation number indicate that dark halos with \( \sim 10^{12} \, h_{70}^{-1} \) Msun, have four galaxies on average, and only about 1/10 (\( \approx 0.3/4 \)) of them have active star formation with a rate exceeding \( \sim 5 \, h_{70}^{-2} \) Msun yr\(^{-1} \) (or \( \sim 20 \, h_{70}^{-2} \) Msun yr\(^{-1} \) if dust extinction is corrected; see Paper V), so that they are identified as LBGs by their bright UV continuum.

Halos hosting SCUBA sources are located near the high-mass end of the dark halo mass function at \( z = 3 \), having \( \sim 3 \times 10^{13} \, h_{70}^{-1} \) Msun. The best estimate of the occupation number of SCUBA sources is about 4, although the uncertainty in it is about an order of magnitude larger than for other objects because of large errors in the measurements in the number density and the correlation length. The large mass of dark halos \( \sim 5 \times 10^{13} \, h_{70}^{-1} \) Msun hosting SCUBA sources indicates that SCUBA sources exist only in protocluster regions.

4.2.3. Dark Halos of LBGs and LAEs at \( z = 5 \)

We next apply our analysis to the LBGs at \( z = 5 \) and LAEs at \( z = 5 \) obtained in Paper II, so as to investigate the evolution of LBGs from \( z = 3 \) to 5 and the differences between LBGs and LAEs. The results are given in rows (14)–(15) of Table 6 and in Figure 11. Both LBGs and LAEs at \( z = 5 \) are embedded in dark halos with \( \sim 10^{12} \, h_{70}^{-1} \) Msun. However, the number densities of LBGs and LAEs are different, and so the occupation number of LAEs \( (N_{\text{occup}} = 3.2) \) is about 5 times larger than that of LBGs \( (N_{\text{occup}} = 0.6) \). It indicates that the majority of dark halos with \( \sim 10^{12} \, h_{70}^{-1} \) Msun do not host LBGs, while hosting multiple LAEs. Since the typical brightness of LAEs in Paper II is about 2 mag fainter in the UV continuum than that of our LBGs at \( z = 5 \), galaxies in the LAE sample have star formation rates (SFRs) about 10 times lower than in the LBG sample. This deep detection limit of the LAEs results in the larger number density for LAEs.

Note that the estimated mass of dark halos hosting LAEs may be smaller than the true values. We use the best-fit power-law function for the angular correlation function shown in Paper II to obtain \( r_0 \) and calculate the galaxy–dark matter bias \( b_0 \) defined at \( r = 8 \, h_{70}^{-1} \) Mpc. However, the angular correlation function has a hump at around \( r = 8 \, h_{70}^{-1} \) Mpc, and the observed points systematically exceed the fitted power law (see Fig. 6 of Paper II). If the hump of the clustering amplitude at \( r = 8 \, h_{70}^{-1} \) Mpc is real, the true \( b_0 \)-value is larger than that obtained in § 3.4, resulting in a larger mass of dark halos \( (3.5^{+1.4}_{-0.7} \times 10^{12} \, h_{70}^{-1} \) Msun) and a larger occupation number (35).

4.3. Descendants of Dark Halos Hosting High-z Galaxies

Once the mass of dark halos is determined at a given redshift (\( z \)), one can calculate analytically the bias of dark halos for their descendants (and progenitors) at an arbitrary redshift. The bias of descendants at the present epoch, \( b_{\text{DH}}^0 \), is calculated by (Sheth et al. 2001)

\[
b_{\text{DH}}^0 = 1 + \frac{D(z)}{\delta_c} \left[ \nu^2 + b_0 \nu^{2(1-c)} - \frac{\nu^{2c}}{\sqrt{a}} \right],
\]

where \( D(z) \) is the growth factor and the other notations are the same as in equation (19). Assuming that the bias thus obtained represents the bias of a given dark-halo mass at the present epoch, we estimate the dark-halo mass of descendants with equation (19). We first calculate the present-day bias for dark halos hosting high-z galaxies and then derive the mass of the descendants at \( z = 0 \) from \( b_{\text{DH}}^0 \).
are for galaxies with redshift. The bias increases monotonically with redshift.

The thick solid line corresponds to the present epoch. It is found from Figure 12 that the majority of descendants are the same as those measured for high-redshift galaxies. In this figure, the number densities of the descendants are the same as those measured for high-redshift galaxies, since the analytic model does not predict the number density of descendants. In other words, we assume that high-redshift galaxies conserve the number densities until the present epoch. It is found from Figure 12 that the majority of high-redshift galaxies are embedded in dark halos of $10^{13}\sim 10^{15} h_7^{\odot} M_\odot$ at the present epoch. Generally speaking, dark-halo masses of $10^{13}\sim 10^{15} h_7^{\odot} M_\odot$ at $z = 0$ correspond to the mass scale of clusters and groups. Our finding means that the high-redshift galaxies examined here become member galaxies of clusters and groups in the present-day universe in a statistical sense. This result is qualitatively consistent with those found by Nagamine (2002), who claims from a hydrodynamic simulation that all bright ($M_V \leq -23$) LBGs at $z = 3$ evolve into galaxies in clusters or groups at $z = 0$ and that half of LBGs with $-23 \leq M_V \leq -21$ fall into clusters or groups.

In Figure 12 the dotted line denotes the number density of present-day galaxies, which is calculated from the occupation number function derived by van den Bosch et al. (2003), who used 2dFGRS data. Note that this occupation number function is for galaxies brighter than $M_b \sim -19$ (this limiting magnitude is determined by the depth of the 2dFGRS data). In other words, the dotted line of Figure 12 indicates the number density of normal bright galaxies. This number density of present-day normal galaxies is found to exceed those of bright LBGs and SCUBA sources. On the other hand, the number density of faint LBGs ($L \leq L^*$), FIRES galaxies, and LAEs is higher than that of present-day normal galaxies. If we assume that all high-redshift galaxies evolve into present-day normal galaxies, the excess of the number density of high-redshift galaxies over that of present-day galaxies puts a constraint on the minimum number of mergers that high-redshift galaxies have experienced until $z = 0$. We define $N_{\text{min}}^\text{merge}$ as the ratio of the number density of high-redshift galaxies to that of present-day galaxies and calculate $N_{\text{min}}^\text{merge}$ for each of the high-redshift galaxy populations in Table 6. It is found that $L \leq L^*$ LBGs, LAEs, and FIRES galaxies should experience mergers at least $2\sim 5$, and 4, and 7 times, respectively, so as to reduce their number density. Part of these high-redshift galaxies will evolve into massive cluster galaxies through mergers. Our findings are consistent with the results of semianalytical models. For example, Governato et al. (2001) find that among 12 halos containing at least one LBG in a protocluster region at $z = 3$, seven halos merge into one central object of the cluster, while the other five halos survive as separate entities inside the cluster at the present epoch.

The thick solid curve in Figure 12 presents the relation between the observed number density of high-redshift galaxies and the predicted dark-halo mass of their descendants at $z = 0$. This curve is derived by evolving the thick solid curve in Figure 9 to the present epoch using equation (20). It is interesting to note that the data points for LBGs at $z = 3$ and 5 are located close to this solid curve. This may imply that descendants of LBGs at $z = 3$, 4, and 5 are indistinguishable in terms of number density and halo mass.

We find that the mass of descendants of dark halos hosting SCUBA sources is about $4.5^{+2.6}_{-1.2} \times 10^{14} h_7^{\odot} M_\odot$. Thus, the descendants of SCUBA sources will probably be embedded in clusters of galaxies. We then find that the data point of SCUBA sources is located near the high-mass end of the thick solid curve in Figure 12. This would indicate that SCUBA sources are very dusty and high-SFR LBGs (a similar discussion is found in Adelberger & Steidel 2000). The number density of SCUBA sources is about 1/10 that of present-day normal galaxies in clusters. This implies that the descendants of SCUBA sources are a rare population in present-day cluster members. Since the fraction of elliptical galaxies in present-day cluster members is about 20% (Dressler 1980; Whitmore et al. 1993), the descendants of SCUBA sources could be cluster ellipticals. This inference supports the prediction by Shu et al. (2001) based on the CDM model that descendants of high-redshift submillimeter sources should reside in present-day clusters and that these objects are likely to be the progenitors of giant ellipticals.

5. DISCUSSION

5.1. Clustering Properties of LBGs at $z = 4$ and 5

Figure 7 shows the observed correlation length, $r_0$ (in comoving units), and galaxy–dark matter bias, $b_0$, of LBGs at $z \approx 4$ and 5 in our sample, together with those of galaxies at various redshifts between $z = 0$ and 3 taken from the literature. The data points for LBGs at $z = 3$–5 are for galaxies with $L \geq L^*$ in the UV continuum, while the other points at $0 < z < 1$ are for galaxies selected by optical continuum. Figure 7 indicates that the clustering of LBGs at $z = 3$–5 is strongly biased against the underlying dark matter (dotted or dashed line). The values for the bias parameter are $b_0 = 3$–5 at $z = 3$–5. The bias increases monotonically with redshift. The increase in $b_0$ with redshift has been predicted by semianalytic models (Baugh et al. 1999; Kauffmann et al.)
and hydrodynamic simulations (Blanton et al. 2000; Yoshikawa et al. 2001). The observed $b_g$-values are consistent with predictions by Kauffmann et al. (1999) (Fig. 7, solid line) for galaxies with $M < -19 + 5 \log h$ in the $B$ band. Although our selection criteria of galaxies (LBG selection) are different from those of Kauffmann et al. (8-band selection), their semianalytic model reproduces the observed trend of $b_g$ fairly well.

In Figure 7, the decrease with redshift over $z \sim 0-1$, has a minimum at $z \sim 1-3$, and then slightly increases at $z > 3$. This behavior can be explained by a combination of (1) the increase in bias ($b_g$) with redshift and (2) the decrease in the correlation amplitude of dark matter ($\xi_{DM}$) with redshift, since $r_0$ is expressed as

$$r_0/(8 \text{ Mpc}) = b_g(2)\xi_{DM}(r = 8 \text{ Mpc})^{1/3}$$

(eq. [15]). At low redshifts, the decrease in $\xi_{DM}$ (with redshift) dominates over the increase in $b_g$. This is the reason why $r_0$ decreases with redshift over $z \sim 0-1$. On the other hand, the increase in $b_g$ is more effective at high redshifts, resulting in an increase in $r_0$ with redshift at $z > 3$. Note that these two factors roughly cancel each other, and thus the change in $r_0$ over $z = 0-5$ is modest compared with the change in $b_g$.

Kauffmann et al. (1999) have found a qualitatively similar trend in the evolution of $r_0$ in their galaxy formation models. They have claimed that a dip in $r_0$ between $z = 0$ and 1 occurs in Λ-dominated flat universes, because structures form early and dark halos hosting L* galaxies are unbiased tracers of the dark matter over this redshift range. To summarize, the observed behaviors of $b_g$ and $r_0$ found in this paper can be regarded as a piece of evidence supporting the biased galaxy formation scenario.

5.2. Dark Matter Properties of High-Redshift Galaxies

It is found from Figures 9 and 10 that the mass of dark halos hosting LBGs at $z = 4-5$ ranges from $10^{10.5}$ to $10^{11}$ $h^{-1}_7$ M$_\odot$. Giavalisco & Dickinson (2001), Bullock et al. (2002), and Moustakas & Somerville (2002) have estimated the mass of dark halos hosting LBGs at $z = 3$ using analytic approaches based on the CDM model. Giavalisco & Dickinson (2001) have found that the average mass of dark halos for LBGs with $R = 23, 25.5$, and 27 is $\langle M \rangle = 3.6, 1.3$, and $0.6 \times 10^{10}$ $h^{-1}_7$ M$_\odot$, respectively. Bullock et al. (2002) have derived the minimum mass of dark halos, $M_{\text{min}}$, and the mass of dark halos hosting one LBG, $M_1$, using the observed spatial correlation functions given by Adelberger et al. (1998), Giavalisco & Dickinson (2001), and Arnouts et al. (1999). They have found $M_{\text{min}} \sim (0.6-1.1) \times 10^{10}$ and $M_1 = (0.9-1.4) \times 10^{10}$ $h^{-1}_7$ M$_\odot$. Moustakas & Somerville (2002) have obtained $M_{\text{min}} \sim 1.9 \times 10^{10}, M_1 \sim 8.6 \times 10^{10}$, and $\langle M \rangle \sim 8 \times 10^{11}$ $h^{-1}_7$ M$_\odot$. From Figure 10, dark halos hosting LBGs at $z = 3$ have $M_1 \sim 10^{13} h^{-1}_7$ M$_\odot$, which is close to the values obtained by Bullock et al. (2002). Similarly, the typical mass of dark halos hosting LBGs at $z = 3$, $\langle M \rangle \sim 10^{13} h^{-1}_7$ M$_\odot$, is close to those derived by Giavalisco & Dickinson (2001) and Moustakas & Somerville (2002). Thus, our results for LBGs at $z = 3$ ($M = (1-10) \times 10^{12}$) are consistent with those given by the previous works. Furthermore, Hamana et al. (2004) have derived recently the mass of dark halos of our LBGs and LAEs at $z = 4$ and 5 using the halo occupation function. They directly fitted predicted angular correlation functions and number density to data for LBGs and LAEs (obtained in this paper and Paper II) and obtained $\langle M \rangle, M_1,$ and $M_{\text{min}}$. They assumed the power-law occupation number function and the same analytic halo models (Sheth & Tormen 1999) as used in this paper. They obtained $\langle M \rangle = 9 \times 10^{11}, 6 \times 10^{11},$ and $1 \times 10^{12}$ $h^{-1}_7$ M$_\odot$ for LBGs at $z = 4$ and 5 and LAEs at $z = 5$, respectively, which is comparable to our results ($M = 4 \times 10^{11}, 9 \times 10^{11},$ and $1 \times 10^{12}$ $h^{-1}_7$ M$_\odot$). Hamana et al. (2004) have found an indication that the simple halo model may not work for small-scale clustering ($\theta < 120$'$''$ corresponding to $r < 3.1$ $h^{-1}_7$ Mpc) of LAEs. In our analysis, we obtain the bias at a large scale (at 8 $h^{-1}_7$ Mpc) and do not examine whether $\xi_{DM}$ is fitted to $\xi(LAE)$ on small scales. In this sense, our analysis is similar to theirs.

In Table 7 we summarize the properties of dark halos hosting $L \geq L^*$ LBGs (and LAEs), which are extracted from Table 6. We find the typical mass of dark halos hosting $L \geq L^*$ LBGs to be $1.3 \times 10^{10}$ to $1.0 \times 9.9 \times 10^{12}$, and $8.9 \times 6.4 \times 10^{11}$ $h^{-1}_7$ M$_\odot$, for $z = 3, 4,$ and 5, respectively. Thus, the masses of dark halos hosting $L \geq L^*$ LBGs are almost constant around $1 \times 10^{12}$ $h^{-1}_7$ M$_\odot$, which is comparable to the total mass of the Milky Way at the present epoch. Although there might be a trend that the typical mass decreases slightly with redshift, we find no clear difference in the mass range of dark halos hosting LBGs with $L > L^*$ ($M \leq -20.5$ or, equivalently, SFR $\geq 10$ and $40 h^{-1}_7 M_\odot$ yr$^{-1}$ before and after extinction correction, respectively) over $z = 2-5$. This finding suggests that the mass-to-luminosity ratio of LBGs (the mass of a hosting halo divided by the UV luminosity of an LBG in the halo) does not largely change from $z = 3$ to 5. In other words, the star formation efficiency is almost constant over $z = 3-5$. This

| Sample Name | $\langle z \rangle$ | Selection Criteria | $\langle M \rangle$ ($h_7 M_\odot$) | $\log(n)$ | $N_{\text{occ}}$ | $\langle M \rangle$ ($z = 0$) ($h_7 M_\odot$) | $N_{\text{min}}$ (merge) |
|-------------|-------------------|---------------------|------------------|-----------|---------------|------------------|------------------|
| UGR-LBG (SPEC) | 3.0 | $R < 25.0$ | $1.3 \times 10^{10}$ | $-3.6 \pm 0.06$ | 0.3 | $7.1 \times 10^{13}$ | 1.1 |
| BR-LBG | 4.0 | $i' < 25.3$ | $1.0 \times 10^{12}$ | $-3.8 \pm 0.05$ | 0.3 | $9.6 \times 10^{13}$ | 0.9 |
| Ve-LBG | 4.7 | $z' < 25.8$ | $8.9 \times 10^{11}$ | $-3.8 \pm 0.3$ | 0.6 | $1.3 \times 10^{14}$ | 0.9 |
| LAE | 4.9 | NB711 $< 25.8$ | $1.1 \times 10^{12}$ | $-3.2 \pm 0.3$ | 3.2 | $1.4 \times 10^{14}$ | 3.8 |

**Notes.**—The quantities $\langle M \rangle$ ($h_7 M_\odot$) and $\langle M \rangle$ ($z = 0$) ($h_7 M_\odot$) are the mass of dark halos hosting high-$z$ galaxies and the typical mass of their descendants at $z = 0$, respectively. $\log(n)$ is the number density of high-$z$ galaxies in units of mag$^{-1}$ $h_7$ Mpc$^{-3}$. $N_{\text{occ}}$ is the occupation number, and $N_{\text{min}}$ (merge) is the minimum number of mergers required from high-$z$ to $z = 0$, assuming that the high-$z$ galaxies evolve into present-day galaxies with $M_6 \leq -19$ mag.

$^a$ $U$-band dropout LBGs given by Giavalisco & Dickinson (2001).
suggestion may not be consistent with the theoretical prediction that the star formation efficiency increases with redshift (e.g., Hernquist & Springel 2003).

Figure 9 and Table 6 show that the occupation number \( N_{\text{ occup}} \) of LBGs for massive dark halos \((\sim 10^{13} h_{70}^{-1} M_\odot)\) is close to unity, while \( N_{\text{ occup}} \) for less massive halos \((\sim 10^{12} h_{70}^{-1} M_\odot)\) is about 0.1. Thus, it is concluded that more massive halos possess a larger number of LBGs. Note that a similar feature is found for the 2dFGRS galaxies at \( z = 0 \) in the range of dark halo mass of \( 10^{12} - 10^{13} h_{70}^3 M_\odot \), as shown in Figure 12.

5.3. Unified View for Various Kinds of High-\( z \) Galaxies

In this section, we discuss the relations between LBGs, LAEs, \( K \)-band–selected galaxies (FIRES), and SCUBA sources. These galaxies are named after the search techniques (\S 1) and are all located at high redshifts. Little has been known to date about the relations between these galaxy populations.

Figure 12 plots the observed number densities of these high-\( z \) galaxies against the predicted masses of dark halos of their descendants at \( z = 0 \). In this figure, the thick solid curve corresponds to the best-fit function to the observed relation between the bias parameter and the number density for our \( z = 4 \) LBGs (Fig. 8 of \S 3.6.1). Interestingly, the data points of LBGs at both \( z = 3 \) and 5 (blue and red circles) are located very close to this curve. This suggests that LBGs at \( z = 3-5 \) are a single population in terms of the mass of hosting halos and the halo occupation number.

We propose here that LBGs at \( z = 3-5 \) come from a single parent population of galaxies, assuming that any dark halo more massive than a critical value has such parent galaxies and that the stellar mass of these galaxies is proportional to the mass of the dark halo. These parent galaxies can be detected by near-infrared surveys such as FIREs, because near-infrared surveys sample galaxies according to their rest-frame optical luminosity, i.e., their stellar mass. Figure 10 shows that the number density of FIREs galaxies is 10 times higher than that of LBGs at the same mass range of dark halos. If LBGs are star-forming galaxies with intermittent star formation activity as suggested in \S 4.2.1, this difference in the number density probably reflects the total duration of active star formation in the galaxies. Both LBGs and FIREs galaxies are seen at \( z = 2.5 \) and 3.5, and the cosmic time between these two redshifts is \( 8.0 \times 10^8 \) yr. If we assume that LBGs are visible in optical observations (i.e., bright in the rest-frame UV wavelength) only when they are in a phase of active star formation, then the total duration of star formation between \( z = 2.5 \) and 3.5 is estimated to be \( 8.0 \times 10^8/10 = 8 \times 10^7 \) yr. This value is comparable to the typical age of stellar populations of LBGs, \( 7 \times 10^7 \) yr, given by Papovich et al. (2001). They have derived the age from a fitting of model spectra to the observed optical to near-infrared colors of LBGs at \( z \sim 3 \), assuming that the star formation history of LBGs is approximated by an exponentially decaying SFR. Furthermore, if we assume that the typical SFR of LBGs is \( 20-200 M_\odot \) yr\(^{-1} \) (whose values are extinction-corrected by a factor of 4: see Fig. 16 and \S 5 of Paper V), then the stellar mass accumulating from \( z = 5 \) to 3 (\( \Delta t = 9.6 \times 10^8 \) yr) is estimated to be roughly

\[
\frac{9.6 \times 10^8 \text{ yr}}{10} (20-200 M_\odot \text{ yr}^{-1}) \sim (2-20) \times 10^9 M_\odot.
\]

This value is comparable to the typical stellar masses of LBGs at \( z \sim 3 \) derived by Shapley et al. (2001) and Papovich et al. (2001). On the other hand, if LBGs are FIRES galaxies in a phase of star formation, LBGs should have stellar masses similar to FIRES galaxies. This prediction is supported by the finding of Franx et al. (2003) that while FIRES galaxies are slightly less luminous in the observed-frame \( K \) band than the brightest LBGs, their stellar masses are comparable to, or higher than, those of the brightest LBGs.

The data point of SCUBA sources (star) is located at the high-mass end of the solid curve in Figure 12. This coincidence would indicate that SCUBA sources and LBGs belong to the same population. As discussed in \S 3.6.1, brighter LBGs have larger dark-halo masses, indicating that LBGs with higher SFRs reside in more massive dark halos. The SFRs of SCUBA sources, \( \sim 10^2 M_\odot \) yr\(^{-1} \), are (much) higher than those of LBGs, 20–200 \( M_\odot \) yr\(^{-1} \) after extinction correction. Thus, the correlation found for our LBGs at \( z = 4 \) holds for SCUBA sources as well. We have also found a possible connection of dusty LBGs to SCUBA sources in the analysis of \S 3.6.2.

We have found a probable connection between LBGs, FIRES galaxies, and SCUBA sources, as discussed above. However, LAEs at \( z = 4.9 \) cannot be understood in a similar way. Since LAEs are also star-forming galaxies, they should be located near the solid curve in Figure 12, if they belong to the same population as LBGs. In this case, LAEs are expected to be located near the low-mass end of the solid curve, because the typical SFR of LAEs is only about 10 and 3 \( M_\odot \) yr\(^{-1} \) with and without dust extinction correction, respectively.\(^{18} \)

These values are comparable to, or lower than, the average extinction-corrected SFR for the faintest subsample of BRi-LBGs, although the dust extinction amount of LAEs is not yet well constrained. However, as found from Figure 12, LAEs deviate largely from the solid curve toward higher halo masses; dark halos hosting LAEs are as massive as those hosting bright LBGs. We think that this apparent discrepancy is related to (1) a large cosmic variance in the LAE sample of Paper II and/or (2) a selection bias of LAEs. First, as we note in \S 3.3, the surveyed volume for their LAEs is just \( 9.0 \times 10^4 h_{70}^{-3} \) Mpc\(^3 \), and thus the results could be strongly affected by the field variance. In reality, Paper IV has found that their LAEs make a large-scale structure, and it is possible that the clustering derived in Paper II is biased high because of this. Second, star-forming galaxies with relatively bright \( \text{Ly}\alpha \) emission can be selected as LAEs even when their SFRs (i.e., UV luminosities) are considerably low. We have found in Paper II that LAEs brighter in \( \text{Ly}\alpha \) emission are clustered more strongly, while clustering strength does not clearly correlate with UV luminosity. If this correlation holds for all star-forming galaxies, the clustering strength of LAEs should be higher than that of general star-forming galaxies with similar SFRs, resulting in very massive dark halos for LAEs. The cause of this “\( \text{Ly}\alpha \) luminosity bias” is not clear to us, but a possible explanation is that the \( \text{Ly}\alpha \) luminosities of galaxies increase (only) when they infall toward massive dark halos. When galaxies infall toward massive halos, star formation may be triggered in them because of, for instance, interactions with the halos. Since the duration of strong \( \text{Ly}\alpha \) emission is very short, galaxies with strong \( \text{Ly}\alpha \) emission will be found only around massive halos, and thus LAEs will have strong

\(^{18} \) We estimate these SFR values from their median UV continuum flux. If we assume LAEs to be dust free, we obtain the typical SFR of about 3 \( M_\odot \) yr\(^{-1} \). On the other hand, if we assume that the dust extinction of LAEs is comparable to that of LBGs at \( z = 4 \), i.e., \( E(B-V) = 0.15 \), we obtain 10 \( M_\odot \) yr\(^{-1} \) for the extinction-corrected SFR.
clustering. Scannapieco & Thacker (2003) have discussed a similar scenario for high-z starburst galaxies. In this explanation, we assume that Ly$\alpha$ luminosity correlates with SFR fairly well. However, this assumption may not necessarily be true. For instance, Shapley et al. (2001) find that LBGs with stronger Ly$\alpha$ emission have older stellar populations, contrary to the predictions from simple H II region models. In any case, our knowledge about the nature of LAEs is still poor. It is of great importance to accumulate data on LAEs by new observations to infer more accurately their fundamental quantities, such as halo mass, stellar mass, and SFR.

Finally, we summarize our unified view for high-z galaxies (i.e., LBGs, FIRES galaxies, SCUBA sources, and LAEs) that we have proposed in this subsection. At $z \sim 3$–5, dark halos with $\sim 10^{12} h_{70}^{-1} M_\odot$ host a few galaxies ($\S$ 4.2.2) whose stellar masses are large enough to be detected as FIRES galaxies. Most of the galaxies hosted by these dark halos are quiescent in star formation, but about 1/10 have moderately strong star formation activity ($\S$ 5.3). These galaxies may be due to the variety of SFRs. These galaxies accumulate stars through episodic star formation activity of $20$–$200 h_{70}^{-2} M_\odot$ yr$^{-1}$ and have stellar masses of $\sim 1 \times 10^{10} M_\odot$ at $z = 3$.

In $\S$ 5.3 we proposed that LBGs and SCUBA sources are high-z star-forming galaxies that have emerged from the same parent population. In our proposal, high-z galaxies experience many episodic star-forming phases, with SFRs of $20$–$200 h_{70}^{-2} M_\odot$ yr$^{-1}$ and whose characteristic total duration is $\sim 80$ Myr (per 1 Gyr). The diversity of the properties of star-forming galaxies may be due to the variety of SFRs. These galaxies accumulate stars through episodic star formation activity of $20$–$200 h_{70}^{-2} M_\odot$ yr$^{-1}$ and have stellar masses of $\sim 1 \times 10^{10} M_\odot$ at $z = 3$.

In Paper V we have found that the LF of LBGs does not change from $z = 4$ to 3, while the number of bright LBGs ($M_{1700} \leq -22$; extinction-corrected SFR is $\geq 100 h_{70}^{-2} M_\odot$ yr$^{-1}$) drops toward $z = 5$. If we assume that the star formation efficiency is constant from $z = 5$ to 4, this implies that massive galaxies (with large cold gas reservoirs) are rare at $z \geq 5$ since matter (baryon + dark matter) has not been assembled by the gravitational instability at $z = 5$. However, the cosmic SFR density is almost constant from $z \sim 5$ to 1 (Paper V). This is because the major contributors to the cosmic star formation are galaxies with smaller SFRs, $\sim 1 h_{70}^{-2} M_\odot$ yr$^{-1}$ (Paper V). Then, most of the galaxies with active star formation of $\sim 100 h_{70}^{-2} M_\odot$ yr$^{-1}$ disappear in the present-day universe (Fig. 16 of Paper V). This reduction of galaxies with intense star-forming activity leads to the decrease in the SFR density from $z \sim 1$ to 0.

On the other hand, dark matter is continuously assembled throughout cosmic time (Fig. 7). Thus, the clustering amplitude, or correlation length, of the dark matter monotonically increases with time. However, the correlation length of galaxies with $L \geq L^*$ does not significantly change from $z = 5$ to 0 because the galaxy–dark matter bias ($b_D$) decreases with time. Thus, the galaxy clustering at $z = 5$ is similar to that at the present epoch. This is the reason why Paper IV discovered the large-scale structure of galaxies at $z = 5$. The size of the large-scale structure at $z = 5$ is comparable to the total mass of the Milky Way. The values of the bias parameter for such massive halos are very high at the observed redshifts. The large masses of dark halos hosting high-z galaxies imply that they will have evolved into groups and clusters of galaxies at the present epoch ($\S$ 4.3 and Fig. 12) in a statistical sense. This would indicate that the majority of present-day field galaxies were formed at $z \geq 3$ or that the high-z progenitors of field galaxies are much fainter than the limiting magnitudes of today’s observations ($i^\prime \sim 27$ and $K \sim 24$).

6. CONCLUSIONS

We detect 2600 Lyman break galaxies (LBGs) at $z = 3.5$–5.2 in the deep ($i^\prime \sim 27$) and wide-field (1200 arcmin$^2$) data in the Subaru Deep Field (SDF) and the Subaru/XMM-Newton Deep Field (SXDF). First, we derive the clustering properties (angular correlation function, correlation length, etc.) of LBGs at $z = 4$ and 5 and present the observational results. We then combine these observational results with the cold dark matter model to understand the formation and evolution of galaxies (luminous matter) and dark halos (dark matter) simultaneously and obtain the theoretical implications. The major findings of our study are categorized into (1) observational results and (2) theoretical implications as follows:

1. Observational results.
   a) We calculate the angular two-point correlation functions of LBGs and find clear signals for LBGs at $z = 4$ and 5. We estimate the correlation length, $r_0$, of the spatial two-point correlation function ($\xi$) by Limber deprojection using the redshift distribution obtained by simulations ($\S$ 2). We obtain $r_0 = 4.1 \pm 0.2$ and $5.9^{+1.3}_{-1.0} h_{70}^{-1}$ Mpc for LBGs at $z = 4$ and 5, respectively. We also calculate $r_0$ for bright galaxies ($L \geq L^*$, i.e., $i^\prime < 25.3$ for LBGs at $z = 4$ and $i^\prime < 25.8$ for LBGs at $z = 5$) to obtain $r_0 = 5.1^{+1.6}_{-1.0}$ and $5.9^{+1.3}_{-1.0} h_{70}^{-1}$ Mpc for LBGs at $z = 4$ and 5 and compare them with that of $z = 3$ LBGs with similar luminosities, to find that the correlation length is almost constant at $\sim 5 h_{100}^{-1}$ Mpc from $z = 3$ to 5.
   b) We find that the correlation amplitude of galaxies is larger than that of the underlying dark matter ($\xi_{DM}$) at high redshifts (at least at $z \geq 3$). We calculate the galactic dark matter bias, $b_D (\xi_{DM})$ by Limber deprojection using the redshift distribution obtained by simulations ($\S$ 2). We find that the $b_D (\xi_{DM})$ of $L \geq L^*$ LBGs at $z = 4$ and 5 is $3.5^{+0.8}_{-0.6}$ and $4.6^{+0.9}_{-1.2}$, respectively, implying that the distribution of LBGs at $z \geq 3$ is highly biased against the underlying dark matter. The bias of galaxies brighter than $\sim L^*$ monotonically increases with redshift, supporting the biased-galaxy formation scenario.
   c) We make subsamples of LBGs at $z = 4$ binned according to magnitude and calculate the correlation length for each subsample. We find that a brighter subsample has a larger correlation length (or larger galactic dark matter bias): the brightest subsample ($M \geq M^* - 1$) has $r_0 = 7.9^{+1.7}_{-2.2} h_{100}^{-1}$ Mpc, while the faintest subsample ($M \geq M^* + 0.5$) has $r_0 = 2.4^{+0.5}_{-0.3} h_{100}^{-1}$ Mpc. Similarly, we make subsamples of LBGs at $z = 4$ according to $E(B-V)$ estimated from the UV-continuum color, $i^\prime - z^\prime$. We find that the dustiest LBGs in our sample
5. It implies that dark halos hosting LBGs are more massive than dark halos existing at $z = 3$, while the number density of FIRES galaxies is 4 times lower than that of dark halos. This implies that about 1/10 ($\approx 0.3/4$) of galaxies residing in dark halos with $\approx 10^{12} h_{70}^{-1} M_\odot$ have active star formation with a rate exceeding $\sim 5 h_{70}^{-2} M_\odot$ yr$^{-1}$ ($\sim 20 h_{70}^{-2} M_\odot$ yr$^{-1}$ if dust extinction is corrected), so that they are identified as LBGs by their bright UV continuum. This difference between LBGs and FIRES galaxies may support the idea that LBGs are galaxies that happen to be bright at UV wavelength because of episodic star formation.

d) We find that the typical mass of dark halos hosting L$\gg$L$^\ast$ LBGs is about $1 \times 10^{12} h_{70}^{-1} M_\odot$, which is comparable to the total mass of the Milky Way at the present epoch, and that the typical mass is almost constant over $z = 3$. It implies that the mass-to-luminosity ratio of LBGs (the mass of a hosting halo divided by the UV luminosity of an LBG in the halo) does not largely change from $z = 3$ to 5 or, equivalently, that the star formation efficiency is almost constant over $z = 3$.

e) Using the CDM model, we estimate the mass of present-day dark halos hosting descendants of high-z ($z = 3$–5) galaxies. We find that the masses of the descendants range from $10^{13}$ to $10^{15} h_{70}^{-1} M_\odot$, which is comparable to the mass of present-day clusters and groups. Thus, dark halos hosting the observed high-z galaxies will evolve into clusters and groups.

In other words, most of the descendants of the high-z galaxies studied here are probably member galaxies in clusters and groups today. The number densities of FIRES galaxies, LAEs, and faint LBGs exceed that of present-day galaxies with $M_b \leq -19$. If most of these high-z galaxies are progenitors of present-day galaxies with $M_b \leq -19$, they should experience mergers at least a few times from $z = 3$ to $z = 0$.

We would like to thank the Subaru Telescope staff for their invaluable help in commissioning the Suprime-Cam that made these difficult observations possible. We thank the anonymous referee for suggestions and comments that improved this article. M. O. acknowledges support from the Japan Society for the Promotion of Science (JSPS) through JSPS Research Fellowships for Young Scientists.

**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

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

Bardeen, J. M., Bond, J. R., Kaiser, N., & Szalay, A. S. 1986, ApJ, 304, 15

Baugh, C. M., Benson, A. J., Cole, S., Frenk, C. S., & Lacey, C. G. 1999, MNRAS, 305, L21

Baugh, C. M., Gardner, J. P., Frenk, C. S., & Sharples, R. M. 1996, MNRAS, 283, L15

Blanton, M., Cen, R., Ostriker, J. P., Strauss, M. A., & Tegmark, M. 2000, ApJ, 531, 1

Bouche, N., & Lowenthal, J. D. 2003, ApJ, 596, 810

Brunner, R. J., Szalay, A. S., & Connolly, A. J. 2000, ApJ, 547, 527

Bullock, J. S., Wechsler, R. H., & Somerville, R. S. 2002, MNRAS, 329, 246

Campos, A., Yahil, A., Windhorst, R. A., Richards, E. A., Pascarellle, S., Impey, C., & Petry, C. 1999, ApJ, 511, L1

Carlberg, R. G., Yee, H. K. C., Morris, S. L., Lin, H., Hall, P. B., Patton, D., Savicki, M., & Shepherd, C. W. 2000, ApJ, 542, 57

Carroll, S. M., Press, W. H., & Turner, E. L. 1992, ARA&A, 30, 499

Chapman, S. C., Blain, A. W., Ivison, R. J., & Smail, I. R. 2003, Nature, 422, 695

Colless, M., et al. 2001, MNRAS, 328, 1039

Connolly, A. J., et al. 2002, ApJ, 579, 42

Daddi, E., Broadhurst, T., Zamorani, G., Cimatti, A., Rottgering, H., & Renzini, A. 2001, A&A, 376, 825

Daddi, E., Cimatti, A., Pozzetti, L., Hoekstra, H., Rettgering, H. J. A., Renzini, A., Zamorani, G., & Mannucci, F. 2000, A&A, 361, 535

Daddi, E., et al. 2003, ApJ, 588, 50

Dodelson, S., et al. 2002, ApJ, 572, 140

Dressler, A. 1980, ApJ, 236, 351

Efstathiou, G., Bernstein, G., Tyson, J. A., Katz, N., & Guhathakurta, P. 1991, ApJ, 380, L47

Franx, M., et al. 2003, ApJ, 587, L79

Fukugita, M., Shimasaku, K., & Ichikawa, T. 1995, PASP, 107, 945

Furusawa, H., Shimasaku, K., Doi, M., & Okamura, S. 2000, ApJ, 534, 624

Fynbo, J. P. U., Ledoux, C., Möller, P., Thomsen, B., & Burud, I. 2003, A&A, 407, 147

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

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
