

## GALAXY CLUSTERING AND GALAXY BIAS IN A $\Lambda$ CDM UNIVERSE

DAVID H. WEINBERG

Department of Astronomy, Ohio State University, 140 West 18th Avenue, Columbus, OH 43210-1173; dhw@astronomy.ohio-state.edu

ROMEEL DAVÉ<sup>1</sup>

Steward Observatory, University of Arizona, 933 Cherry Avenue, Tucson, AZ 85721-0065; rad@as.arizona.edu

NEAL KATZ

Department of Astronomy, University of Massachusetts, 710 North Pleasant, Amherst, MA 01003-9305; nsk@kaka.astro.umass.edu

AND

LARS HERNQUIST

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA, 02138; lars@cfa.harvard.edu

Received 2002 December 16; accepted 2003 October 6

### ABSTRACT

We investigate galaxy clustering and the correlations between galaxies and mass in the  $\Lambda$ CDM cosmological model (inflationary cold dark matter with  $\Omega_m = 0.4$ ,  $\Omega_\Lambda = 0.6$ ,  $h = 0.65$ ,  $n = 0.95$ , and  $\sigma_8 = 0.8$ ), using a large, smoothed particle hydrodynamics simulation (SPH; with  $2 \times 144^3$  particles in a  $50 h^{-1}$  Mpc cube). Simulated galaxies can be unambiguously identified as clumps of stars and cold gas a few kpc to a few tens of kpc across, residing in extended halos of hot gas and dark matter; the space density of the resolved galaxy population at  $z = 0$  corresponds to that of observed galaxies with luminosity  $L \gtrsim L_*/4$ . We investigate the galaxy correlation function, the pairwise velocity dispersion and mean pairwise velocity, and the second and third moments of counts in cells; we also investigate the galaxy-mass correlation function and the average extended mass distributions around galaxies, both of which can be measured via galaxy-galaxy lensing. For the most part, the predicted biases between galaxies and dark matter lead to good agreement with current observations, including (1) a nearly constant comoving correlation length from  $z = 3$  to 0 for mass-selected galaxy samples of constant comoving space density; (2) an rms bias factor  $b_\sigma \approx 1$  at  $z = 0$ ; (3) a scale-dependent bias on small scales that transforms the curved dark matter correlation function into a nearly power-law galaxy correlation function; (4) galaxy pairwise dispersion and hierarchical skewness ratio  $S_3$  in good agreement with observed values, lower than values for the dark matter by  $\sim 20\%$ ; (5) a ratio of galaxy-galaxy to galaxy-mass correlation functions consistent with recent measurements from the Red Cluster Sequence survey; and (6) a mean excess mass  $\Delta M(260 h^{-1} \text{ kpc})$  approximately proportional to galaxy baryon mass  $M_b$ , in agreement with estimates from the Sloan Digital Sky Survey (SDSS). All these clustering properties vary with galaxy baryon mass and, more strongly, with the age of a galaxy's stellar population. The predicted dependences are in good qualitative agreement with the observed dependence of galaxy clustering and the galaxy-mass correlation function on galaxy type. The predicted ratio  $\Delta M(260 h^{-1} \text{ kpc})/M_b$  is lower than the SDSS estimates by a factor of  $\sim 1.5\text{--}3$  for galaxies with  $M_b \gtrsim 2 \times 10^{11} M_\odot$ . A test with a higher resolution (smaller volume) simulation suggests that this discrepancy is largely a numerical artifact; if so, then the SDSS weak-lensing comparison leaves limited room for feedback or other astrophysical processes to reduce the stellar masses of luminous galaxies, at least given our adopted cosmological parameters. On the whole, our results show that the  $\Lambda$ CDM model and the galaxy formation physics incorporated in the SPH simulation give a good account of observed galaxy clustering, but anticipated improvements in clustering and weak-lensing measurements will soon test this picture in much greater detail.

*Subject headings:* cosmology: theory — galaxies: formation — large-scale structure of universe — methods: numerical

### 1. INTRODUCTION

The clustering of galaxies has long been an essential testing ground for cosmological models and the theory of galaxy formation, with comparisons between predicted and observed clustering driving much of the progress in cosmology during the 1970s and 1980s. Such comparisons provided early support for gravitational instability as the central mechanism of structure formation (e.g., Davis, Groth, & Peebles 1977), for initial conditions with a power spectrum redder than white

noise (e.g., Gott & Rees 1975; Gott & Turner 1977), for a “bottom up” rather than “top down” sequence of gravitational clustering (e.g., White, Frenk, & Davis 1983; Davis et al. 1985; Fry & Melott 1985), for an approximately Gaussian distribution of primordial fluctuations (e.g., Gott et al. 1989; Weinberg & Cole 1992; Bouchet et al. 1993), and for a power spectrum significantly redder than that predicted by the “standard” ( $\Omega_m = 1$ ,  $h \equiv H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1} = 0.5$ ) cold dark matter scenario (e.g., Efstathiou et al. 1990; Maddox et al. 1990; Park et al. 1994). The growth of galaxy redshift surveys has led to measurements of increasing precision and detail, and cosmological  $N$ -body simulations have developed into a

<sup>1</sup> Hubble Fellow.

powerful tool for calculating the gravitational clustering of collisionless dark matter from specified initial conditions. The main obstacle to drawing stronger inferences from the data is the dependence of theoretical predictions on *both* the relatively straightforward physics of gravitational clustering, which largely determines the distribution of dark matter, and the more complex physics of galaxy formation, which determines the relation between galaxies and mass, often referred to generically as “bias.”

In this paper, we study galaxy clustering and galaxy bias in a large ( $2 \times 144^3$  particles,  $50 h^{-1}$  Mpc cube), smoothed particle hydrodynamics (SPH) simulation of a low-density inflationary cold dark matter universe with a cosmological constant ( $\Lambda$ CDM). Our goals are, first, to see whether the currently leading cosmological model and “standard” galaxy formation physics as incarnated in our SPH simulation can reproduce existing observations and, second, to give guidance for the physical interpretation of those observations and for the design of more demanding tests of the theoretical predictions. The  $N$ -body+hydro approach to ab initio predictions of galaxy bias has a substantial history, beginning with dissipative “sticky particle” simulations by Carlberg & Couchman (1989) and continuing in the early 1990s with simulations using SPH (Katz, Hernquist, & Weinberg 1992; Evrard, Summers, & Davis 1994) or Eulerian grid hydrodynamics (Cen & Ostriker 1992) to model galaxy formation in a representative (but small) cosmological volume.

The present paper is a direct descendant of Katz et al. (1992), but we study a different cosmological model, we incorporate improvements in the physical treatment of radiative cooling and star formation (Katz, Weinberg, & Hernquist 1996, hereafter KWH), and, above all, we take advantage of the parallel implementation of TreeSPH (Davé, Dubinski, & Hernquist 1997) and advances in computer technology to simulate a volume more than 90 times larger than that of Katz et al. (1992), at similar resolution. Our analysis overlaps significantly with other recent studies using large-volume SPH (Pearce et al. 1999, 2001; Yoshikawa et al. 2001) or Eulerian hydrodynamics (Cen & Ostriker 2000) simulations. Relative to these investigations, we have higher mass and/or spatial resolution and a somewhat smaller simulation volume, as discussed in § 2 below.

Two other approaches to ab initio predictions of bias have gained currency in recent years: high-resolution collisionless  $N$ -body simulations that identify galaxies with “subhalos” in the dark matter distribution (e.g., Colín et al. 1999; Kravtsov & Klypin 1999) and a hybrid method that combines  $N$ -body simulations of the dark matter component with semianalytic treatments of the galaxy formation physics (e.g., Kauffmann, Nusser, & Steinmetz 1997; Governato et al. 1998; Kauffman et al. 1999a, 1999b; Benson et al. 2000a, 2000b, 2001; Hatton et al. 2003). We discuss comparisons of our results to those from other hydrodynamic simulations and from the high-resolution  $N$ -body and hybrid approaches as they arise. We carry out a detailed comparison between our simulation and the semianalytic model of Benson et al. (2000a) in a separate paper that focuses on the “halo occupation distribution” (HOD) predicted by the two methods (Berlind et al. 2003). The HOD description can be used to calculate many different galaxy clustering statistics, and it helps explain the origin of bias in a physically intuitive manner (see Berlind & Weinberg 2002 and references therein). On the basis of this comparison, we conclude that our SPH approach and the semianalytic

method of Benson et al. (2000a) should yield similar predictions for most galaxy clustering statistics.

Here we focus mainly on the “classic” measurements of galaxy clustering—the two-point correlation function, variance of galaxy counts, and pairwise velocity moments—and on one of the simplest measures of higher order clustering, the third moment of counts in cells. Large redshift surveys such as the 2dF Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS) now allow precise measurements of these quantities for multiple classes of galaxies, defined by luminosity, color, morphology, or spectral type (see, e.g., Norberg et al. 2001, 2002b; Zehavi et al. 2002 and references therein). While we cannot model these variations with galaxy type in detail—our simulation volume is far smaller than these surveys, and we do not resolve the morphology of the simulated galaxies—we can examine the predicted trends of clustering with baryon mass, stellar population age, and local environment. We concentrate on present-day clustering, but we also compute the evolution of the two-point function, which can be compared with results from deep redshift surveys and to studies of Lyman break galaxies at  $z \approx 3$  (e.g., Adelberger et al. 1998, 2003). The discovery that the clustering of Lyman break galaxies is similar to that of  $L_*$  galaxies at  $z = 0$ , despite the weaker expected clustering of the underlying mass distribution, provides strong evidence that the bright galaxy population was highly biased at  $z \approx 3$ , even if galaxies roughly trace mass today.

Galaxy-galaxy weak-lensing is an important new observational probe of the relation between galaxies and dark matter, measuring the galaxy-mass cross-correlation function and the extended mass distributions around galaxies of different types and in different environments. We devote considerable attention to modeling recent observations of this phenomenon. The closest similar efforts are those of White, Hernquist, & Springel (2001), who present predictions from a hydrodynamic simulation at  $z = 1$  and 0.5, and Guzik & Seljak (2001) and Yang et al. (2003), who analyze the  $N$ -body+semianalytic galaxy distribution of Kauffman et al. (1999a).

One strength of the hydrodynamic simulation approach is that it predicts properties of the intergalactic medium (IGM) in addition to the galaxy population. We have used the simulation analyzed here to model the phase distribution of baryons in the present-day universe (Davé et al. 2001), the X-ray background (Croft et al. 2001), X-ray emission from galaxy groups (Davé, Katz, & Weinberg 2002), and X-ray absorption by the diffuse IGM (Chen et al. 2003). In combination with other simulations, we have used it to study cooling radiation and submillimeter emission from young galaxies (Fardal et al. 2001, 2002), the relative importance of mergers and smooth accretion in galaxy assembly (Murali et al. 2002), and the correlations between galaxies and the Ly $\alpha$  forest at high redshift (Kollmeier et al. 2003). The observed distributions and correlations of galaxy properties, in particular the luminosity function and the Tully-Fisher (Tully & Fisher 1977) relation, are also essential tests of the galaxy formation model, and any discrepancies with these observations can provide insight into the model’s failings. We consider these characteristics of the galaxy population in a separate paper (M. A. Fardal et al. 2004, in preparation), although the comparison to observed galaxy-galaxy lensing measurements brings up some of the same issues here. We proceed to a description of our numerical methods in § 2, to predictions of galaxy clustering in § 3, and to the galaxy-mass correlation function and comparison

with lensing measurements in § 4. We recapitulate our findings and discuss their implications in § 5.

## 2. SIMULATING GALAXY FORMATION

### 2.1. Simulation

We employ a hydrodynamic simulation of a  $\Lambda$ CDM model with  $\Omega_m = 0.4$ ,  $\Omega_\Lambda = 0.6$ ,  $h = 0.65$ ,  $\Omega_b = 0.02 h^{-2} = 0.0473$ , inflationary spectral index  $n = 0.95$ , and a power spectrum normalization  $\sigma_8 = 0.80$ . This model is consistent with most available observational constraints (Bahcall et al. 1999; Jaffe et al. 2001), including COBE normalization (Bennett et al. 1996), cluster masses (Eke, Cole, & Frenk 1996), cosmic shear (Hoekstra, Yee, & Gladders 2002 and references therein), the deuterium abundance in high-redshift Lyman limit systems (Tytler, Fan, & Burles 1996; Burles & Tytler 1997, 1998), the Hubble diagram of Type Ia supernovae (Riess et al. 1998; Perlmutter et al. 1999), and the flux power spectrum of the Ly $\alpha$  forest (Croft et al. 1999, 2002; McDonald et al. 2000). Results from the *Wilkinson Microwave Anisotropy Probe* experiment imply a somewhat lower value of  $\Omega_m = 0.33 \pm 0.05$  for our adopted value of  $h$  (Spergel et al. 2003), as does a joint analysis of the 2dFGRS and earlier cosmic microwave background data (Percival et al. 2002). At the time that we carried out our simulation, the constraint on  $\sigma_8 \Omega_m^{0.5}$  implied by the cluster X-ray temperature function (Eke et al. 1996) and the low  $\sigma_8$  favored by the Ly $\alpha$  forest (Croft et al. 2002) led us to choose  $\Omega_m = 0.4$ , but some recent analyses suggest that cluster abundances are consistent with and perhaps favor a lower value of  $\sigma_8 \Omega_m^{0.5}$  (e.g., Seljak 2002b; Bahcall et al. 2003). We expect that the primary effect of a lower  $\Omega_m$  on our results would be to reduce galaxy peculiar velocities, roughly in proportion to  $\Omega_m^{0.6}$ , and to reduce the total-to-baryon mass ratios computed in § 4, roughly in proportion to  $\Omega_m$ .

Our initial conditions represent a random  $50 h^{-1}$  Mpc (comoving) cube, realized with  $144^3$  gas and  $144^3$  dark matter particles in a triply periodic volume, yielding particle masses of  $m_{\text{gas}} = 8.5 \times 10^8 M_\odot$  and  $m_{\text{dark}} = 6.3 \times 10^9 M_\odot$ , respectively. These initial conditions are evolved from redshift  $z = 49 \rightarrow 0$  by using Parallel TreeSPH (Davé et al. 1997). We soften the gravitational force via a  $10 h^{-1}$  kpc (comoving) spline kernel (see the appendix to Hernquist & Katz 1989), approximately equivalent to a Plummer softening<sup>2</sup> of  $7 h^{-1}$  kpc. For particle time steps we use the criteria described by KWH and Quinn et al. (1997), setting the tolerance parameter  $\eta$  (defined in these papers) to 0.4. The simulation was evolved to  $z = 0.5$  on the Cray T3E at San Diego Supercomputing Center and finished on the SGI Origin 2000 at National Center for Supercomputing Applications.

We include radiative cooling for primordial composition gas as described by KWH. The tests in Weinberg, Hernquist, & Katz (1997) show that photoheating by an ionizing background artificially suppresses the formation of low-mass galaxies in a simulation of this resolution, but that a simulation evolved without an ionizing background produces much the same galaxy population (above its resolution limit) as a higher resolution simulation with an ionizing background. Although Weinberg et al. (1997) considered an  $\Omega_m = 1$  CDM

model rather than  $\Lambda$ CDM, we expect their conclusions to carry over to this model, and we therefore evolve the simulation with no ionizing background.

We incorporate star formation by using the algorithm described and tested by KWH. This algorithm converts cold dense gas into collisionless stars at a rate governed by the local dynamical and cooling timescales, returning supernova feedback as thermal energy to the surrounding medium. The tests in KWH and Weinberg et al. (1997) show that the resulting galaxy population is insensitive to the one free parameter of the algorithm and is similar to that identified using cold gas in a simulation without star formation; gravity, gasdynamics, and radiative cooling determine where galaxies form and how massive they will be. A radical change to the star formation or feedback algorithm might alter the galaxy population in a more significant way. However, we use only the rank order of galaxy masses and median stellar population ages to define the galaxy samples in this paper, not the absolute values, so our clustering results would be unaffected by any change in the star formation and feedback algorithms that preserved this rank order. Only the total-to-baryon mass ratios discussed in § 4 (Fig. 14 in particular) are sensitive to the absolute values of the galaxy baryon masses.

Our numerical approach is similar to those used by Pearce et al. (1999, 2001) and Yoshikawa et al. (2001), but we have opted for a somewhat smaller volume and larger particle number, and consequently higher resolution. Pearce et al. use  $128^3$  particles in a  $70 h^{-1}$  Mpc cube, so our volume per particle ratio is higher by  $(70/50)^3 (144/128)^3 = 3.9$ . We consider this the most useful measure of mass resolution because it tracks the cutoff of small-scale power in the initial conditions; however, the ratio of SPH particle masses is only 2.3 because Pearce et al. use a lower value of  $\Omega_b$  and a higher value of  $h$ , and this may be the more relevant figure for cooling physics. The gravitational softening of the Pearce et al. simulation is similar to ours at  $z = 0$  ( $10 h^{-1}$  Plummer vs. our  $\sim 7 h^{-1}$  kpc) but larger at higher redshift, since they keep the softening fixed in physical coordinates at  $z < 2.5$ . The mass resolution of the Yoshikawa et al. (2001) simulation is similar to that of Pearce et al. ( $128^3$  particles in a  $75 h^{-1}$  Mpc cube), and the (Plummer equivalent) gravitational softening is  $41 h^{-1}$  kpc, constant in comoving coordinates. The most important methodological difference is that we convert cold gas into collisionless stars, while Pearce et al. and Yoshikawa et al. “decouple” cold gas from hot gas when computing SPH densities but leave the cold gas subject to SPH forces and dissipation. Despite the differences in numerical resolution, input physics, and implementation details, we find quite good agreement with these previous investigations where our results overlap, as discussed in § 3. The cosmological parameters of the three  $\Lambda$ CDM simulations are similar but not identical. (Pearce et al. investigate an  $\Omega_m = 1$  model in addition to  $\Lambda$ CDM.)

Cen & Ostriker (2000) have investigated galaxy clustering and bias in the  $\Lambda$ CDM model with a  $512^3$  Eulerian mesh simulation of a  $100 h^{-1}$  Mpc volume. The initial mass resolution of this simulation is higher than ours, but the force resolution is much lower, with individual mesh cells  $\approx 200 h^{-1}$  kpc. Because the simulation does not track individual galaxies within high-density regions, Cen & Ostriker examine stellar mass-weighted clustering statistics rather than the galaxy number-weighted statistics that we investigate here. Nonetheless, our conclusions about galaxy bias on the scales  $r \gtrsim 1 h^{-1}$  Mpc resolved by the Cen & Ostriker simulation are qualitatively similar, as discussed in § 3.

<sup>2</sup> The spline and Plummer forms are, of course, different, so there is no exact transformation of softening lengths. The spline-softened force is larger for  $r > 0.75\epsilon$ , and it reaches 70% of the Newtonian force at  $r = 1.07\epsilon$ , 0.7 times the radius at which the Plummer force reaches the same strength. The spline-softened force is Newtonian at  $r > 2\epsilon$ , while the Plummer force only approaches the Newtonian force asymptotically.

## 2.2. Galaxy Identification

Figure 1 illustrates the state of the simulation at  $z = 0$ , showing dark matter, gas, and star particles projected through the entire volume in a quarter of the box, and in expanded views 6, 1.5, and  $0.375 h^{-1}$  Mpc on a side centered on the most massive system in our volume. The star particles are clumped into groups of extremely high overdensity, from a few kpc to a few tens of kpc across. Dark halos with the mass scale of galaxy groups and clusters contain many such dense clumps in addition to a smooth distribution of hot gas. Figure 1 shows that there is virtually no ambiguity in identifying the galaxies in this simulation; one simply needs an automated algorithm that picks out the distinct clumps of stars and cold gas. The algorithm that we use is Spline Kernel Interpolative DENMAX<sup>3</sup>

<sup>3</sup> See <http://www-hpcc.astro.washington.edu/tools/skid.html>.

(SKID), which selects gravitationally bound groups of particles that are associated with a common local density maximum (see KWH). SKID selects essentially the same population of objects that one would pick out by eye from a representation such as Figure 1.

Figure 2 shows the baryonic mass function of the simulated galaxy population at  $z = 3, 1,$  and  $0$ , where  $M_b$  represents the total mass of a galaxy's stars and associated cold ( $T < 30,000$  K) gas. The number of galaxies steadily increases with time, although the rate of new galaxies forming is dramatically slower for  $z = 1 \rightarrow 0$  than for  $z = 3 \rightarrow 1$ . We discuss the  $z = 0$  luminosity function and Tully-Fisher relation in a separate paper (M. A. Fardal et al. 2004, in preparation), and we have elsewhere compared predictions of the  $z = 3$  luminosity function from higher resolution simulations (of smaller volumes) with measurements of the Lyman break galaxy population

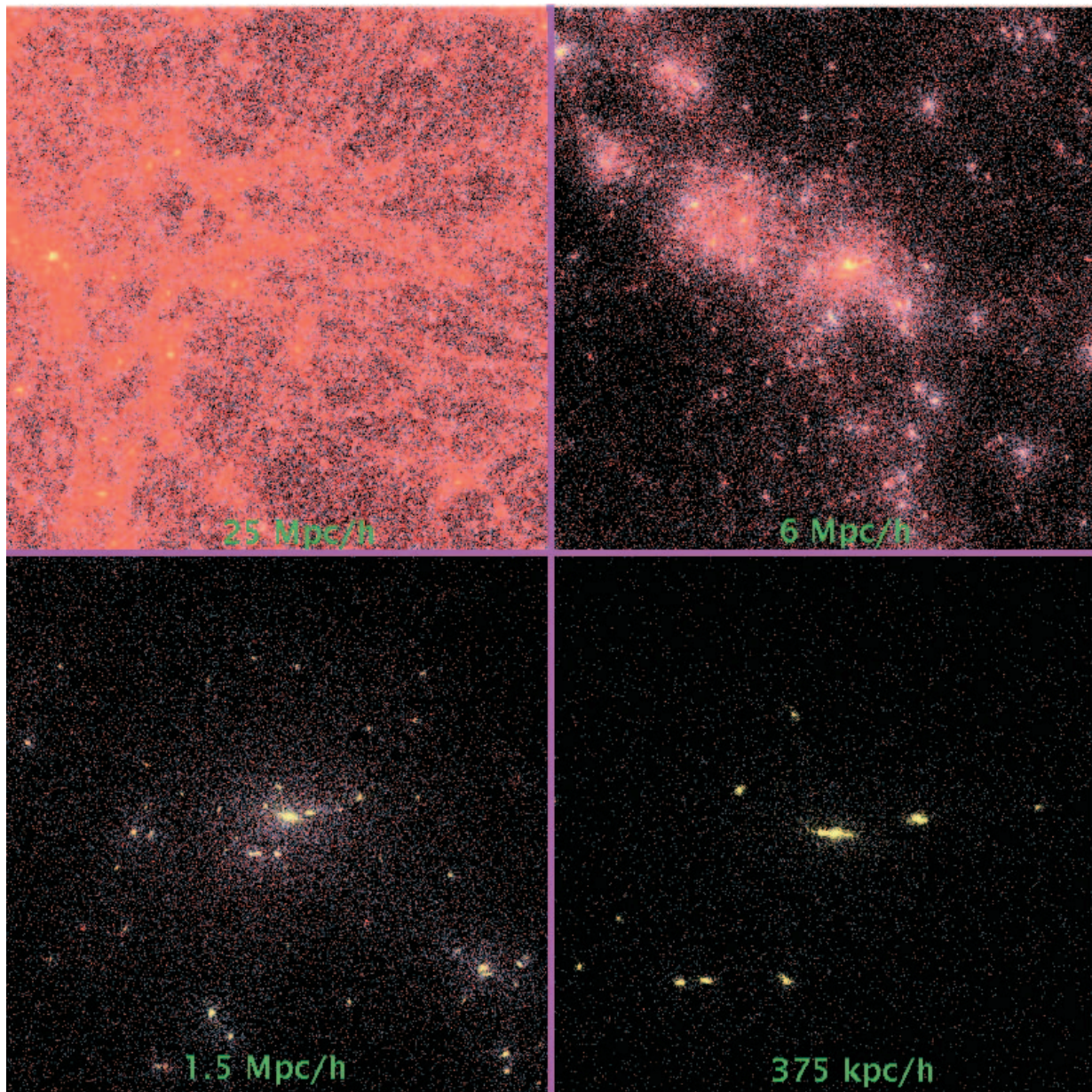


FIG. 1.—Projections through the simulation volume at  $z = 0$ . Blue particles are dark matter, red are gas, and yellow are baryonic particles where star formation has occurred. The size of each view is indicated. The expanded views are centered on the largest galaxy, seen near the left edge of the upper left panel. All panels show the projection through the entire simulation volume.

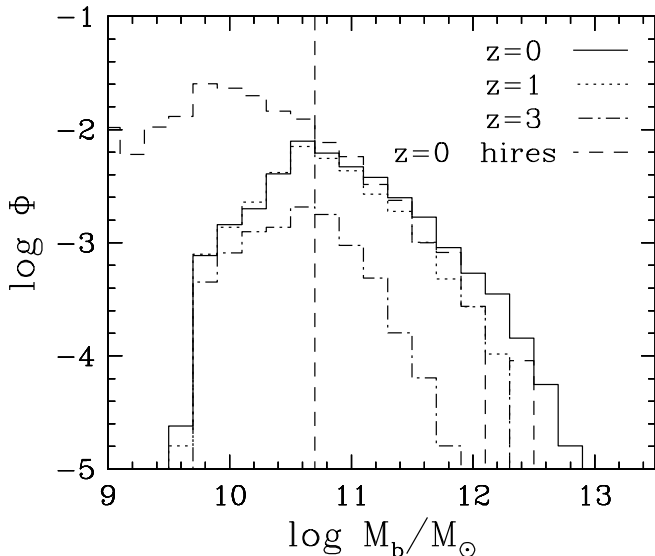


FIG. 2.—Baryonic mass function of SKID-identified galaxies at redshifts  $z = 0$  (solid histogram),  $z = 1$  (dashed histogram), and  $z = 3$  (dot-dashed histogram) in the  $(50 h^{-1} \text{ Mpc})^3$  simulation, in number per  $(h^{-1} \text{ Mpc})^3$  per decade of mass. The dashed histogram shows the  $z = 0$  mass function from a higher resolution simulation of a smaller volume. Solid and dotted histograms diverge below  $M_b = 5 \times 10^{10} M_\odot$  (vertical dashed line), corresponding to  $60m_{\text{gas}}$  in the large-volume simulation, which we adopt as our threshold for identifying resolved galaxies.

(Weinberg et al. 1999; Weinberg, Hernquist, & Katz 2002). The dashed vertical line in Figure 2 marks the mass of 60 SPH particles, and all the simulated mass functions turn over shortly below this threshold. Our tests using a suite of simulations of different resolution show that galaxies below this threshold have underestimated masses or may be missed entirely, while the masses and locations of galaxies above this threshold are generally robust (see Weinberg et al. 1999). The dashed line shows the mass function at  $z = 0$  from a higher resolution ( $\epsilon = 3 h^{-1} \text{ kpc}$ ,  $m_{\text{gas}} = 1.04 \times 10^8 M_\odot$ ), smaller volume ( $22.22 h^{-1} \text{ Mpc}$  cube) simulation having  $2 \times 128^3$  particles and the same cosmological parameters, analyzed in the same way; a comparison with the solid line clearly shows that the turnover below the demarcated mass limit is due to numerical effects. We therefore adopt a baryon mass of  $M_b = 60m_{\text{gas}} = 5 \times 10^{10} M_\odot$  as our completeness limit, and we analyze the clustering of galaxies only above this threshold. Note that the mass function comparison also suggests that the lower resolution simulation may overestimate the baryon masses of the largest galaxies, a point that we return to in § 4.

Of 4632 total SKID-identified galaxies at  $z = 0$ , 2571 satisfy our  $M_b \geq 60m_{\text{gas}}$  completeness criterion. At  $z = 1, 2, 3$ , and 4, the corresponding fractions are 2011/4012, 1034/2488, 420/1177, and 104/384. The sample of all galaxies with  $M_b \geq 60m_{\text{gas}}$  will be referred to as our “complete sample” of galaxies. At  $z = 0$ , the space density of the complete sample,  $2571/(50 h^{-1} \text{ Mpc})^3 = 0.02 h^3 \text{ Mpc}^{-3}$ , corresponds to that of observed galaxies brighter than about  $L_*/4$  (Blanton et al. 2001), where  $L_*$  is the characteristic luminosity of the Schechter (1976) function fit.

### 3. GALAXY CLUSTERING

Figure 3 illustrates the evolution of structure in a  $10 h^{-1} \text{ Mpc}$  thick slice of the simulation volume, running from  $z = 3$  (top) to  $z = 1$  (middle) to  $z = 0$  (bottom). The left column shows the

dark matter distribution, which exhibits the usual behavior seen in  $N$ -body simulations: a steady increase of contrast between growing filaments and emptying voids, and fragmentation of filaments into increasingly massive clumps. The middle column shows the clustering of a subset of galaxies with nearly constant comoving space density, namely, the 420 resolved galaxies at  $z = 3$  and the 500 most massive galaxies at  $z = 1$  and 0. We refer to this subset below as our “ $L_*$  sample.” The evolution of galaxy clustering is markedly different from that of the dark matter. The network of filaments and voids is already present at  $z = 3$ , with nearly the full contrast that it achieves by  $z = 0$ . The principal change in the clustering of this sample is on small scales, with the collapse of some extended structures into tight clumps.

The right column of Figure 3 shows the full resolved galaxy population at the three redshifts, with a point size that reflects the galaxy baryon mass. Here one can see the increase in the number of resolved galaxies (from 420 to 2011 to 2571), the growth in the average mass of these galaxies, and the preferential location of the most massive galaxies in dense systems at  $z = 0$ . The changing number density and size of points makes it difficult to gauge the evolution of clustering strength visually, but we see below that the correlation function of this complete galaxy sample stays nearly constant from  $z = 3$  to  $z = 0$ .

#### 3.1. Two-Point Correlation Function

The two-point correlation function is the best-studied measure of galaxy clustering, in part because it is relatively straightforward to estimate from angular or redshift survey data, in part because it contains complete statistical information on fluctuations in the linear regime if these fluctuations are Gaussian, and in part because its observed form is remarkably simple. For most galaxy samples, the correlation function is well described by a power law,

$$\xi(r) = \left( \frac{r}{r_0} \right)^{-\gamma}, \quad (1)$$

on scales  $r \sim 0.01\text{--}10 h^{-1} \text{ Mpc}$  (Totsuji & Kihara 1969; Peebles 1974; Gott & Turner 1979). However, the values of  $r_0$  and, to a lesser extent,  $\gamma$  depend on the luminosity, color, and type of the sample galaxies (e.g., Guzzo et al. 1997; Norberg et al. 2001, 2002b; Zehavi et al. 2002 and numerous references therein), and high-precision measurements show subtle but statistically significant departures from a power law (Zehavi et al. 2003; Hawkins et al. 2004). We list determinations of  $r_0$  and  $\gamma$  from a number of recent redshift surveys in Table 1. While the varying values of  $r_0$  and  $\gamma$  partly reflect statistical uncertainties, they arise mainly from the different selection criteria of the samples from which they are measured. In most cases, the values quoted in Table 1 are derived from flux-limited galaxy samples, and the weighting of galaxies of different luminosities depends on the specific procedure adopted for estimating  $\xi(r)$ . The quoted values for the 2dFGRS are derived from a volume-limited sample of galaxies with absolute magnitudes close to the characteristic magnitude  $M_{b1}^*$  obtained for 2dFGRS galaxies. The quoted values for the SDSS are from a flux-limited sample, but they are similar to those obtained for a volume-limited sample with absolute magnitudes close to  $M_{b1}^*$ . The difference between the 2dFGRS and SDSS correlation lengths,  $4.9 \pm 0.3$  versus  $6.1 \pm 0.2 h^{-1} \text{ Mpc}$  (or  $6.3 \pm 0.8 h^{-1} \text{ Mpc}$  for the volume-limited sample)

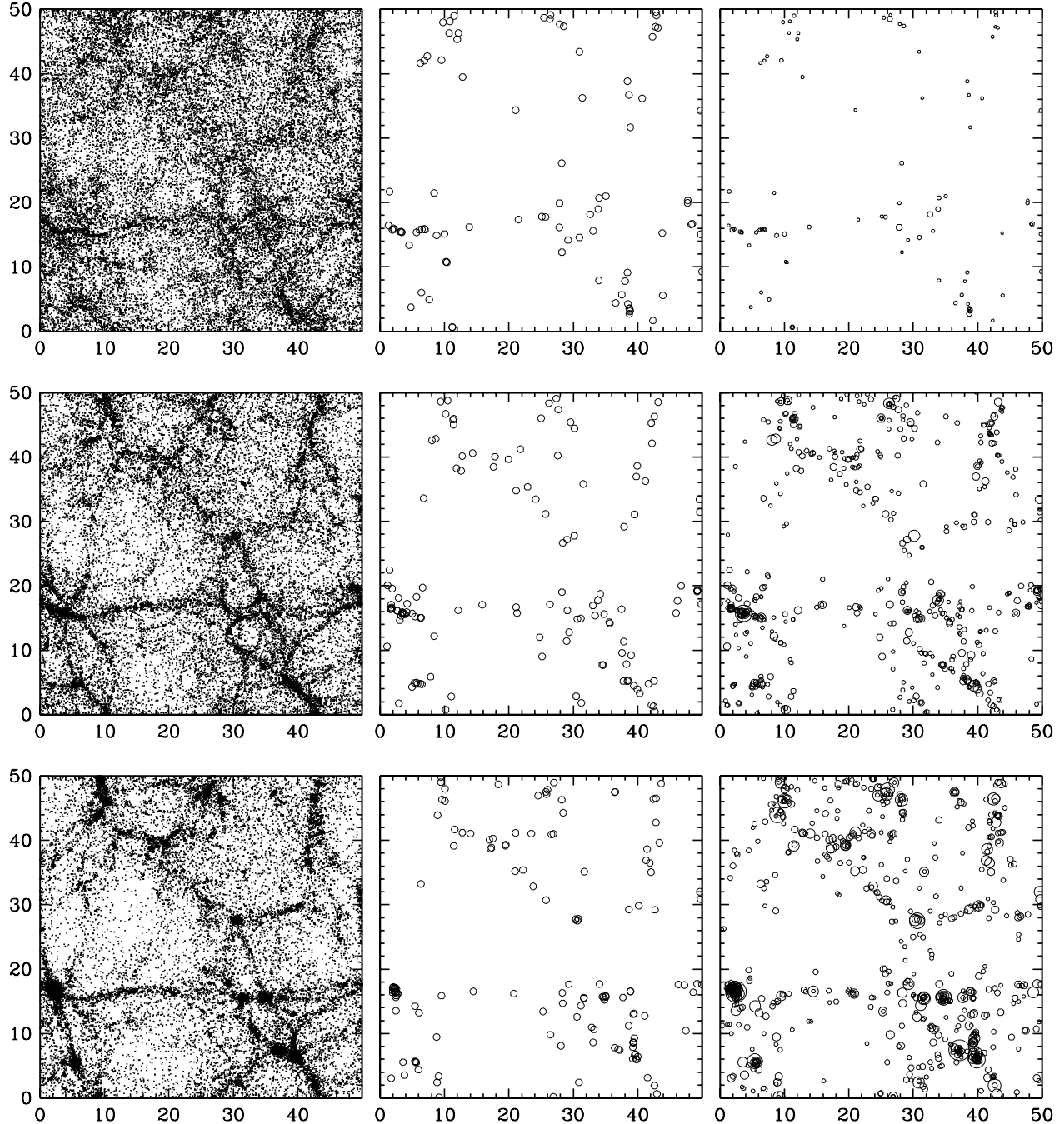


FIG. 3.—Structure in a  $10 h^{-1}$  Mpc thick slice of the  $50 h^{-1}$  Mpc cube at  $z = 3$  (top),  $z = 1$  (middle), and  $z = 0$  (bottom). Left column, randomly selected 5% of the dark matter particles; middle column, the 420 resolved galaxies at  $z = 3$  and the 500 most massive galaxies at  $z = 1$  and 0; right column, all resolved galaxies at the three redshifts, with the area of each point proportional to the galaxy baryon mass and the smallest points corresponding to our resolution threshold of  $5 \times 10^{10} M_{\odot}$ .

presumably reflects the difference between blue and red selection.

Early  $N$ -body simulations showed that gravitational clustering produces an approximate power law  $\xi(r)$ , especially if the initial power spectrum has a slope  $n \approx -1$  (Gott, Turner, & Aarseth 1979). However, the much higher precision correlation functions calculated with modern simulations show significant departures from a power law  $\xi(r)$  for CDM or power-law initial spectra (e.g., Jenkins et al. 1998 and references therein). The origin of these departures can be understood in analytic models of matter clustering (Hamilton et al. 1991; Peacock & Dodds 1996; Ma & Fry 2000; Seljak 2000; Scoccimarro et al. 2001).

Thus, the observed power-law form of  $\xi(r)$  requires either special features in the primordial power spectrum or scale-dependent bias between the galaxy and dark matter correlation functions at the present day.

Figure 4 illustrates the evolution of  $\xi(r)$  in our simulation from  $z = 3$  to 0. Figure 4 (top) shows results for the complete galaxy sample, i.e., all galaxies above our  $60m_{\text{gas}} = 5 \times 10^{10} M_{\odot}$  resolution threshold at the corresponding redshift, corresponding to the right column of Figure 3. As noted in § 2.2, there are 420, 1034, 2011, and 2571 galaxies satisfying this criterion at  $z = 3, 2, 1,$  and 0, respectively. Since the strength of galaxy clustering depends on mass (see § 3.2 below), it is useful

TABLE 1  
REAL-SPACE CORRELATION FUNCTION PARAMETERS  $r_0$  AND  $\gamma$  FROM OUR SIMULATION AND RECENT REDSHIFT SURVEYS

Survey	$r_0$	$\gamma$	Reference
All simulated.....	$3.96 \pm 0.30$	$1.78 \pm 0.05$	
$L_*$ Simulated .....	$4.52 \pm 0.38$	$2.00 \pm 0.08$	
200 most massive .....	$5.20 \pm 0.52$	$2.05 \pm 0.09$	
Stromlo-APM.....	$5.1 \pm 0.2$	$1.71 \pm 0.05$	1
CfA2/SSRS2 <sup>a</sup> .....	5.8	1.8	2
LCRS.....	$5.06 \pm 0.12$	$1.86 \pm 0.034$	3
Durham/UKST.....	$5.1 \pm 0.3$	$1.6 \pm 0.1$	4
ESP.....	$4.15 \pm 0.2$	$1.67 \pm 0.08$	5
2dFGRS <sup>b</sup> .....	$4.9 \pm 0.3$	$1.71 \pm 0.06$	6
SDSS <sup>c</sup> .....	$6.1 \pm 0.2$	$1.75 \pm 0.03$	7
PSCz <sup>d</sup> .....	3.7	1.69	8

<sup>a</sup> Based on Fig. 3 of the Marzke et al. 1995 analysis of CfA2 and SSRS2.

<sup>b</sup> We quote values for the volume-limited subsample with  $-20.5 < M_b < -19.5$ , close to the value of  $L_*$  found by Norberg et al. 2002a for the 2dFGRS.

<sup>c</sup> We quote values for the full flux-limited sample. For a volume-limited sample with  $-21.5 < M_r < -20.0$ , close to the Blanton et al. 2001 value of  $L_*$  for the SDSS, Zehavi et al. 2002 find  $r_0 = 6.3 \pm 0.8$  and  $\gamma = 1.80 \pm 0.09$ .

<sup>d</sup> Note that these are IRAS-selected galaxies and therefore preferentially late types.

REFERENCES.—(1) Loveday et al. 1995; (2) Marzke et al. 1995; (3) Jing et al. 1998; (4) Ratcliffe et al. 1998; (5) Guzzo et al. 2000; (6) Norberg et al. 2001; (7) Zehavi et al. 2002; (8) Jing et al. 2002.

to identify samples of specified space density for comparison with observations. Figure 4 (*middle*) shows results for the 500 most massive galaxies present at each redshift (except at  $z = 3$ , which contains only the 420 galaxies above the resolution threshold). The ranking of each galaxy is based on its total mass of stars and cold ( $T < 30,000$  K) gas. The comoving space density of this sample,  $500/(50 h^{-1} \text{ Mpc})^3 = 0.004 h^3 \text{ Mpc}^{-3}$ , is similar to that of  $L_*$  galaxies today (Blanton et al. 2001) and similar to (but slightly higher than) the space density of Lyman break galaxies in the spectroscopic samples of Adelberger et al. (1998). We therefore refer to this as our “ $L_*$  sample,” illustrated by the middle column of Figure 3. At  $z = 0$ , the baryon mass threshold for this sample is  $M_b = 2.9 \times 10^{11} M_\odot$ . Figure 4 (*bottom*) shows the dark matter  $\xi(r)$ .

In each panel of Figure 4, we show as a reference the power-law fit to the SDSS correlation function found by Zehavi et al. (2002). Comparison of our predictions with results from a flux-selected observational sample rests on the implicit assumption that the luminosity of a galaxy is an approximately monotonic function of its mass, and we use the SDSS result in preference to the 2dFGRS result of Norberg et al. (2001) because this assumption seems more applicable to the  $r$ -band selection used in the SDSS. Lyman break galaxy samples at  $z \approx 3$  are selected in the rest-frame UV, where the instantaneous star formation rate is more important than the stellar mass in determining galaxy luminosity; however, in our simulations, these star formation rates are fairly well correlated (although not perfectly so) with baryon mass (Weinberg et al. 2002).

The main source of statistical error in our predictions is the finite number of coherent structures present in our  $(50 h^{-1} \text{ Mpc})^3$  simulation volume. We estimate the error bars on the correlation function and on other statistical quantities that we compute later by applying jackknife resampling to the eight octants of the simulation cube. Specifically, we create eight jackknife subsamples by deleting each of the eight octants of the cube in turn, and we estimate  $\xi(r)$  for each subsample; to avoid boundary effects that would complicate  $\xi(r)$  estimation and reduce our effective volume, we identify pairs by using the full periodic cube and assign each pair to an octant

on the basis of the position of one of its particles, selected at random. The  $1 \sigma$  error bar  $\sigma_i$  assigned to the value of  $\xi(r)$  at  $r = r_i$  is

$$\sigma_i^2 = \frac{N}{N-1} \sum_{j=1}^N [\xi^j(r_i) - \hat{\xi}(r_i)]^2, \quad (2)$$

where  $\xi^j$  is the estimate in subsample  $j$ ,  $\hat{\xi}$  is the estimate from the full cube, and  $N = 8$ . This approach is similar to using the dispersion among the values of  $\xi(r)$  from the eight octants (in which case one would divide by  $8^{1/2}$  to get the error on the mean), but it is more robust because each of the subsamples is closer in size to the full sample. Zehavi et al. (2002) use this method to estimate statistical errors on the observed correlation functions of SDSS galaxies. The jackknife error estimates automatically incorporate the contribution from galaxy shot noise in addition to the finite number of structures, but the latter usually dominates by a large factor over the former (which would be represented by “Poisson error bars”).

The striking difference between the evolution of galaxy and dark matter correlation functions in Figure 4 is similar to that found in previous studies based on hydrodynamic simulations, semianalytic calculations of galaxy formation, and high-resolution  $N$ -body simulations that identify galaxies as “subhalos” within larger virialized objects (Baugh et al. 1999; Katz, Hernquist, & Weinberg 1999; Colin et al. 1999; Kauffman et al. 1999b; Pearce et al. 1999, 2001; Cen & Ostriker 2000; Benson et al. 2001; Somerville et al. 2001; Yoshikawa et al. 2001). As Figure 3 shows, galaxies (at least those massive enough to be resolved by our simulation) form at special locations in the density field, and at high redshift they already trace out the network of filaments nascent in the dark matter distribution. The dark matter correlation function grows in time as mass moves into this network from the surrounding regions, but the structure traced by galaxies stays relatively unchanged, and the galaxy correlation function is only weakly dependent on redshift.

Figure 5 quantifies these points in terms of the parameters  $r_0$  and  $\gamma$ , which we determine from power-law fits to the

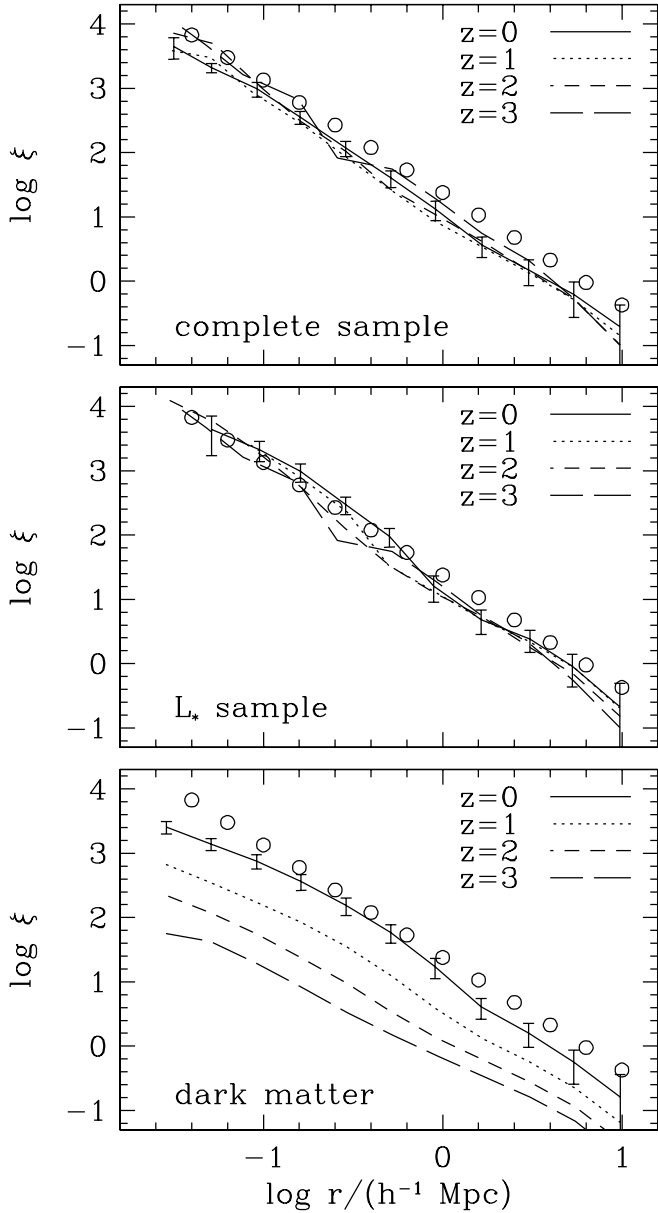


FIG. 4.—Correlation functions  $\xi(r)$  for the complete galaxy sample (*top*), the  $L_*$  galaxy sample (*middle*), and the dark matter (*bottom*), at redshifts  $z = 0, 1, 2,$  and  $3$ . Jackknife error bars are plotted for the  $z = 0$  correlation function. Open circles show a power law with the parameters measured from the SDSS by Zehavi et al. (2002),  $r_0 = 6.1 h^{-1}$  Mpc and  $\gamma = 1.75$ .

correlation functions over the full range plotted in Figure 4,  $0.02 h^{-1}$  Mpc  $< r < 12 h^{-1}$  Mpc. (Note that our figures plot the point for each radial separation bin at the computed mean separation of pairs in that bin.) For some sparse high-redshift samples, the smallest radial bins have no pairs, in which case we start the fit at slightly larger  $r$ . We use the jackknife error bars on the  $\xi(r)$  points to weight the fit and determine its statistical errors. We do *not* incorporate the covariance of the  $\xi(r)$  errors; in principle, we could estimate the full error covariance matrix by the jackknife method, but in practice such estimates are noisy and not easy to use. Treating the  $\xi(r)$  errors as independent causes us to *underestimate* the statistical uncertainty in  $r_0$ . The dark matter correlation functions are not well described by a power law at  $z \lesssim 1$ , where the fits have  $\chi^2$  per degree of freedom (dof) of  $\sim 1.5$ – $2$ , so the derived values

of  $r_0$  and  $\gamma$  are sensitive to the radial range selected for the fit. Power-law fits to the galaxy correlation functions are nearly always adequate, and they usually have  $\chi^2/\text{dof} < 1$  ( $\sim 0.3$ – $0.7$ ) because we have ignored the covariance of the errors in  $\xi(r)$ .

Figure 5 shows steady growth of the comoving correlation length of the dark matter, as expected. The correlation length of the complete galaxy sample, on the other hand, starts at  $r_0 = 5.2 h^{-1}$  Mpc at  $z = 4$ , declines to a minimum of  $3.0 h^{-1}$  Mpc at  $z = 1.5$ , and climbs again to  $r_0 = 4 h^{-1}$  Mpc by  $z = 0$ . The  $L_*$  sample has a higher correlation length than the complete sample at  $z = 0$ ,  $r_0 \approx 4.5 h^{-1}$  Mpc, although the correlation lengths of the two samples converge at higher redshift as the  $L_*$  threshold approaches our completeness threshold. Correlation lengths for a sample with a higher mass threshold, the 200 most massive galaxies at each redshift, are higher still,  $r_0 = 5.2 h^{-1}$  Mpc at  $z = 0$ , with only a slight decline toward higher redshift. The values of  $\gamma$  for the galaxy correlation functions range from 1.75 to 2.05, with a tendency for larger  $\gamma$  for the more massive galaxy samples or at higher redshifts. At high redshift, the dark matter correlation function is close to a power law, but a quite shallow one. As the scale of nonlinear structure grows, it produces a “bulge” in the dark matter  $\xi(r)$  (see Fig. 4), which drives the correlation function away from a power law and steepens the effective value of  $\gamma$ .

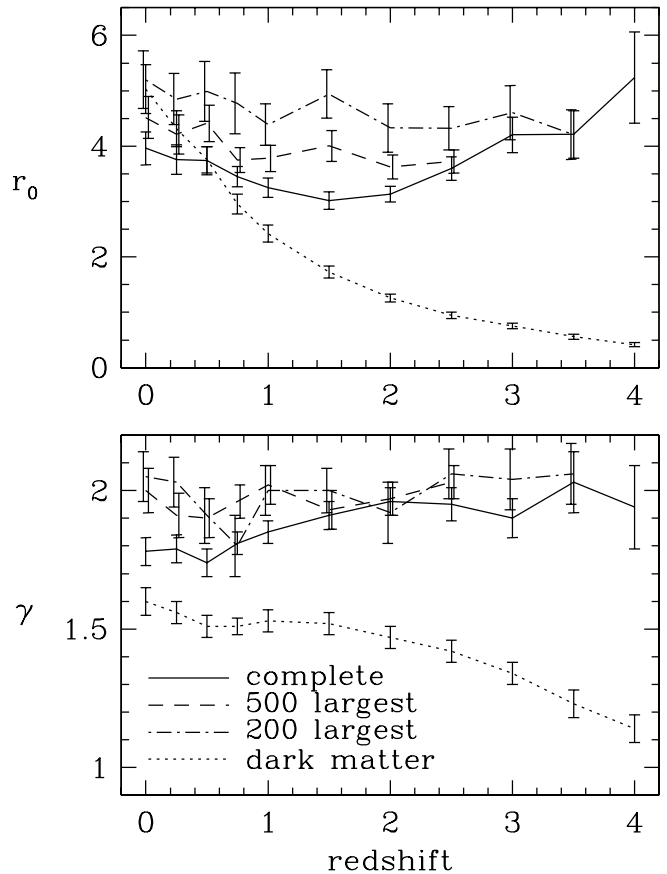


FIG. 5.—Evolution of the correlation length  $r_0$  (in comoving  $h^{-1}$  Mpc) and power-law index  $\gamma$ , for all galaxies (*solid line*), the 500 most massive galaxies (*dashed line*), the 200 most massive galaxies (*dot-dashed line*), and the dark matter (*dotted line*). Error bars are obtained from the power-law fits, using the jackknife errors on  $\xi(r)$ . Lines for the 500 largest galaxies stop at  $z = 2.5$ , since the complete sample contains fewer than 500 galaxies at higher redshift; likewise, lines for the 200 largest galaxies stop at  $z = 3.5$ .



The slow evolution of the galaxy correlation length in Figure 5 agrees with results from the KPNO wide-area survey (Postman et al. 1998), which yields  $r_0 \approx 4.5 h^{-1}$  Mpc (comoving) at  $z \approx 0.5$ , and from the CNOC-2 redshift survey, which shows little evolution of the comoving  $r_0$  out to  $z \sim 1$  (Carlberg et al. 1999). However, the value of  $r_0$  depends significantly on galaxy mass and age (see Fig. 7 below), so caution is required when interpreting results of surveys that may select different types of galaxies at different redshifts. The predicted correlation length at  $z = 3$ ,  $r_0 = 4.2 \pm 0.3 h^{-1}$  Mpc, agrees very well with the value  $r_0 = 4.3 \pm 0.3 h^{-1}$  Mpc measured from spectroscopic samples of Lyman break galaxies (Adelberger et al. 2003; we have transformed from their  $\Omega_m = 0.3$  cosmology to our  $\Omega_m = 0.4$  cosmology by using the ratio of angular diameter distances at  $z = 3$ ), although the slope  $\gamma = 1.90 \pm 0.07$  is steeper than the observed value of  $\gamma = 1.55 \pm 0.15$ . Similar values of  $r_0$  for Lyman break galaxies have been obtained using semianalytic galaxy formation models (Governato et al. 1998; Kauffman et al. 1999b; Benson et al. 2001; Somerville et al. 2001) and other hydrodynamic simulations (Cen & Ostriker 2000; Pearce et al. 2001; Yoshikawa et al. 2001). The value of  $r_0$  from our  $50 h^{-1}$  Mpc simulation is somewhat larger than the estimate  $r_0 = 3.1 h^{-1}$  Mpc (including a factor 2.2 correction for finite box size effects) derived by Katz et al. (1999) from a simulation of an  $11.111 h^{-1}$  Mpc cube, but a difference of this sort is expected because the mass resolution threshold of our present simulation is a factor of 8 higher, and correlation strength increases with galaxy mass.

Perhaps the most significant result of this section is the power-law form of the galaxy correlation function at  $z = 0$ . The dark matter correlation function shows the curvature found in previous  $N$ -body studies (e.g., Jenkins et al. 1998), and it is inconsistent with a pure power law. We demonstrate this point explicitly in Figure 6, in which we plot the ratio of the dark matter, complete galaxy, and  $L_*$ -galaxy correlation functions

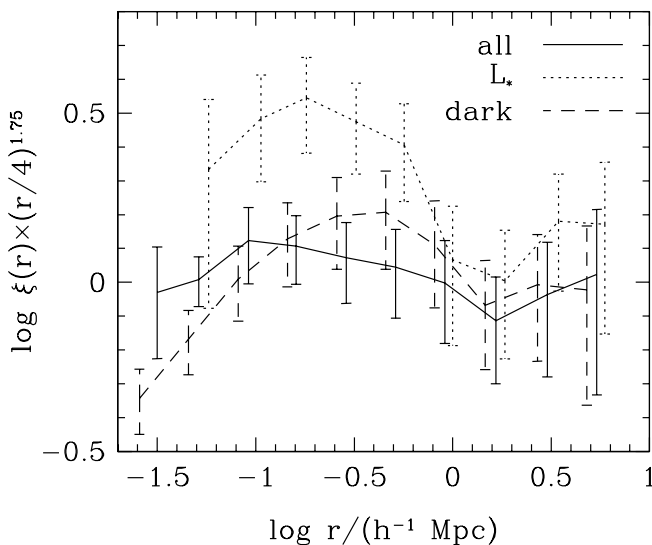


FIG. 6.—Ratio of the  $z = 0$  correlation functions to a power law  $\xi(r) = (r/4.0 h^{-1} \text{ Mpc})^{-1.75}$ , for the complete galaxy sample (solid line), the  $L_*$  sample (dotted line), and the dark matter (dashed line). Small horizontal offsets have been added to the jackknife error bars to improve clarity. On this plot, the SDSS correlation function would be a horizontal line at  $\log(6.1/4.0)^{1.75} = 0.32$ , and the 2dFGRS correlation function of  $L_*$  galaxies would be a horizontal line at  $\log(4.9/4.0)^{1.75} = 0.15$ .

with a power law  $(r/4.0 h^{-1} \text{ Mpc})^{-1.75}$  that has the observed logarithmic slope. The dark matter correlation function rises above this power law at  $r \sim 0.5 h^{-1}$  Mpc and falls well below it at  $r \lesssim 0.05 h^{-1}$  Mpc. (The departure from a power law is somewhat less prominent than that in the  $\Lambda$ CDM simulation of Jenkins et al. 1998 because we adopt  $\sigma_8 = 0.8$  and  $n = 0.95$  instead of  $\sigma_8 = 0.9$  and  $n = 1$ .) The correlation function of the complete sample, on the other hand, follows the power law almost perfectly. The situation for the  $L_*$  sample is less clear; here  $\xi(r)$  shows a rise at  $r \lesssim 1 h^{-1}$  Mpc reminiscent of that shown by the dark matter, but within our statistical errors the  $L_*$   $\xi(r)$  is adequately described by a power law with  $\gamma = 2.00 \pm 0.08$ .

The correlation length of the  $L_*$  sample,  $r_0 = 4.5 \pm 0.4 h^{-1}$  Mpc, is smaller than the SDSS value of  $r_0 = 6.1 h^{-1}$  Mpc and slightly below the value  $r_0 = 4.9 h^{-1}$  Mpc found for blue-selected  $L_*$  galaxies in the 2dFGRS. The slope  $\gamma = 2.00 \pm 0.08$  is steeper than the observed value of  $\gamma \approx 1.75$ . The low amplitude appears to be partly an artifact of the particular realization of  $\Lambda$ CDM initial conditions in our  $50 h^{-1}$  Mpc simulation volume. We have run particle mesh  $N$ -body simulations (using the code of Park 1990) of these initial conditions and of four equivalent volumes with different random realizations, and we find that the dark matter correlation length of this realization is a factor of 1.2 lower than that derived from the average correlation function of the other four realizations ( $3.79$  vs.  $4.60 h^{-1}$  Mpc) simply by chance. (The amplitude of  $\xi$  is systematically low by about the level of the  $1 \sigma$  jackknife error bar at all  $r$ ; as noted earlier, our treatment of the errors as independent in the power-law fits causes the quoted error bar on  $r_0$  to be too small.) There is also a systematic effect of missing power on scales larger than  $50 h^{-1}$  Mpc. With PM simulations of  $100 h^{-1}$  Mpc volumes, we find a further increase in the mean  $r_0$  by a factor of 1.1 and a slightly shallower slope ( $\gamma = 1.92$  over the range  $1$ – $10 h^{-1}$  Mpc vs.  $\gamma = 1.97$  for the  $50$  Mpc realization of the SPH simulation). Unfortunately, we do not know just how these differences in the dark matter  $\xi(r)$  translate to the galaxy  $\xi(r)$ , but the statistical and systematic effects in our  $50 h^{-1}$  Mpc volume are large enough that we do not presently regard the quantitative discrepancies between the predicted and observed  $(r_0, \gamma)$  for  $L_*$  galaxies as significant. The statistical uncertainties can be reduced by using the halo occupation distribution derived from the SPH simulation to populate a larger volume  $N$ -body simulation, as discussed by Berlind et al. (2003).

Berlind & Weinberg (2002) show that the key requirements for obtaining a power-law  $\xi(r)$  in a  $\Lambda$ CDM model are reducing the efficiency of galaxy formation in the most massive (group and cluster scale) halos and suppressing pair counts in the lowest mass halos by keeping fluctuations about the mean occupation number well below the Poisson level (see also Jing, Mo, & Börner 1998; Benson et al. 2000b; Seljak 2000; Peacock & Smith 2000; Scoccimarro et al. 2001). The galaxy population in our simulation satisfies both of these requirements, a point that we discuss in detail elsewhere (Berlind et al. 2003). Our conclusion that including galaxy formation physics removes most of the discrepancy between the predicted correlation function of dark matter and the observed correlation function of galaxies agrees with conclusions derived from other studies based on semianalytic modeling (Kauffman et al. 1999a; Benson et al. 2000a, 2000b; Somerville et al. 2001), high-resolution  $N$ -body simulations

(Colin et al. 1999; Kravtsov et al. 2003), and numerical hydrodynamics (Pearce et al. 1999, 2001; Cen & Ostriker 2000; Yoshikawa et al. 2001).

### 3.2. Correlation Functions of Subpopulations

As mentioned in § 3.1, the galaxy correlation function is known to vary significantly with galaxy luminosity, color, and morphology (e.g., Davis & Geller 1976; Hamilton 1988; Loveday et al. 1995; Guzzo et al. 1997; Willmer, da Costa, & Pellegrini 1998). The most comprehensive recent examinations are the studies of luminosity and color dependence for SDSS galaxies by Zehavi et al. (2002) and the studies of luminosity and spectral type dependence for 2dFGRS galaxies by Norberg et al. (2001, 2002b). Zehavi et al. (2002) find a steady increase of correlation amplitude with luminosity, going from  $\sim 1.5$  mag fainter than  $M_*$  to  $\sim 1.5$  mag brighter, and they find that red galaxies have a higher amplitude and steeper correlation function than blue galaxies, in agreement with earlier studies of the difference between clustering of early-type (red) and late-type (blue) samples. Norberg et al. (2001) find a somewhat different luminosity dependence for the  $b_J$ -selected 2dFGRS galaxies, with little dependence of clustering amplitude on luminosity below  $L_*$  but a strong dependence at higher luminosities. Norberg et al. (2002b) find that the correlation amplitude varies strongly with galaxy spectral type, analogous to the usual dependence on morphological type, and that the luminosity dependence is present within each spectral type class. Galaxies with early spectral types have a somewhat steeper correlation function than those with late spectral types, although the difference in slope is not as striking as that for the SDSS red and blue galaxy samples.

Our simulation volume is too small and our computation of galaxy properties too simple to allow detailed comparisons with these observations, but we can examine qualitative trends. The color and spectral type of a galaxy depend largely on the age of its stellar population, and morphology is also known to correlate strongly with population age. For our simulated galaxy population, we define a galaxy's age by the epoch at which it formed half its stars. While this approach necessarily brings our heuristic star formation algorithm into play, the details of this algorithm should have little effect on the conclusions presented here, since we use only the *relative* ages of the simulated galaxies.

Figure 7a shows the correlation functions of the older (*dashed line*) and younger (*dotted line*) halves of the complete galaxy sample, compared with the correlation function of the complete sample (*solid line*). The correlation amplitude exhibits a clear dependence on galaxy age; the older galaxies have a correlation length of  $r_0 = 5.2 h^{-1}$  Mpc and the younger galaxies have  $r_0 = 2.8 h^{-1}$  Mpc. The fitted slopes of the correlation functions are nearly identical,  $\gamma = 1.84$  and  $1.82$ , respectively. (Values of  $r_0$  and  $\gamma$  for all subpopulations appear in Table 2.) These results are in good qualitative agreement with the spectral type dependence found by 2dFGRS; relative to the SDSS results, they reproduce the observed color dependence of correlation amplitude but not the observed change of slope.

Figure 7b compares the correlation functions of the more massive (*dashed line*) and less massive (*dotted line*) halves of the complete sample. While the more massive galaxies have a higher  $\xi(r)$  at all separations, the difference is marked only below  $\sim 0.5 h^{-1}$  Mpc. Power-law fits yield nearly identical correlation lengths,  $r_0 \approx 4.0 h^{-1}$  Mpc, but significantly

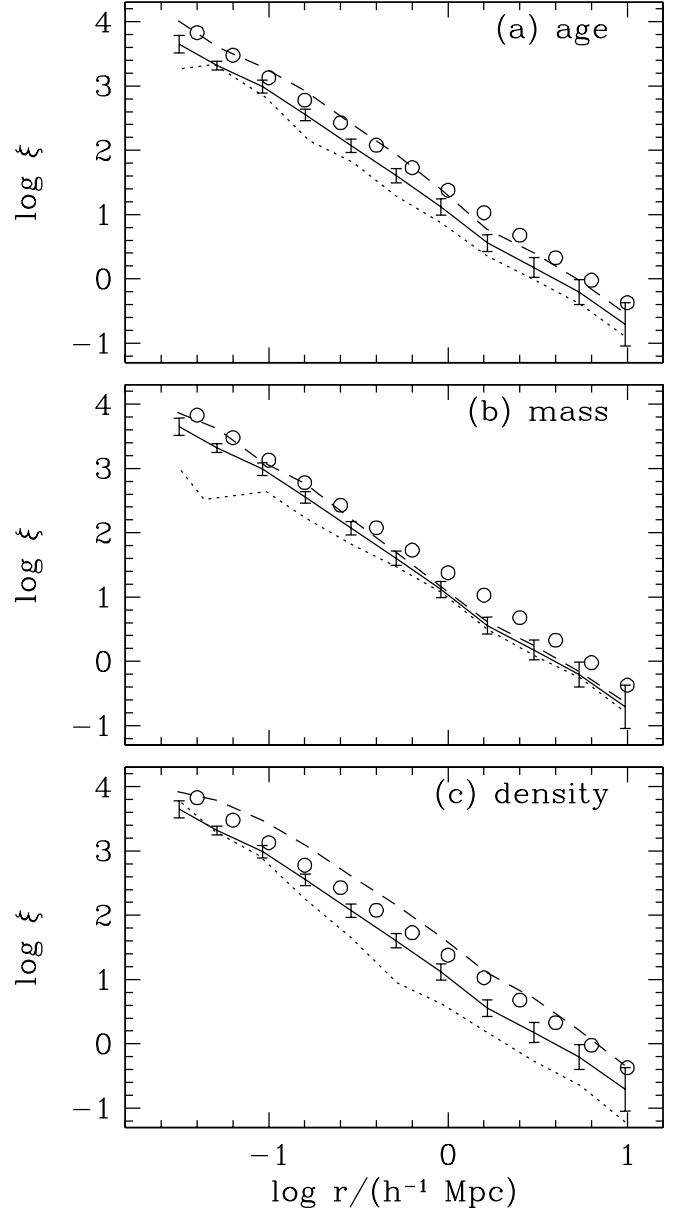


FIG. 7.—Dependence of the galaxy correlation function on (a) stellar population age, (b) baryon mass, and (c) density of the local environment at  $z = 0$ , showing the correlation function and jackknife error bars of the complete galaxy sample (*solid line*) and the power-law fit to the SDSS correlation function (*open circles*). Dashed lines show the correlation function of (a) the older, (b) more massive, and (c) higher density half of the complete sample. Dotted lines in each panel show the correlation function of the other half of the sample.

different slopes ( $\gamma = 1.90$  vs.  $\gamma = 1.57$ ). Table 1 shows that  $r_0$  does rise significantly if one takes the 500 most massive ( $r_0 = 4.5 h^{-1}$  Mpc) or 200 most massive ( $r_0 = 5.2 h^{-1}$  Mpc) galaxies, and in combination with Figure 7b it suggests that this mass dependence sets in at roughly the space density of  $L_*$  galaxies. The 2dFGRS results show a transition of this sort (Norberg et al. 2001), although the SDSS results show a steadier dependence of clustering amplitude on luminosity (Zehavi et al. 2002).

Figure 7c shows the effect of dividing the galaxy sample on the basis of local density, computed by smoothing with a spline kernel (the same form used in the SPH calculation) whose radius is adjusted to always enclose 16 galaxies within

TABLE 2  
 $r_0$  AND  $\gamma$  FROM SIMULATED SUBSAMPLES DECOMPOSED BY AGE, MASS, AND LOCAL GALAXY DENSITY

Subsample	$r_0$	$\gamma$
All .....	$3.96 \pm 0.30$	$1.78 \pm 0.05$
Old.....	$5.24 \pm 0.47$	$1.84 \pm 0.06$
Young.....	$2.77 \pm 0.17$	$1.82 \pm 0.06$
High mass .....	$4.00 \pm 0.29$	$1.90 \pm 0.06$
Low mass.....	$3.95 \pm 0.41$	$1.57 \pm 0.08$
High density.....	$7.82 \pm 0.71$	$1.77 \pm 0.06$
Low density .....	$1.89 \pm 0.08$	$2.07 \pm 0.05$

the smoothing volume. Since  $\xi(r)$  is itself a measure of overdensity at radius  $r$ , it is no surprise that this division produces the most marked change in the correlation function. The enhancement of  $\xi(r)$  persists to scales much larger than the typical smoothing radius because of the bias of locally dense regions toward overdense large-scale environments (Kaiser 1984; see Mann, Peacock, & Heavens 1998 and Narayanan, Berlind, & Weinberg 2000 for discussions in the context of  $N$ -body models). The observed dependence of galaxy morphology on local density (e.g., Dressler 1980; Postman & Geller 1984) should therefore lead to a large-scale bias in the clustering of early-type galaxies.

To summarize the results of this and the previous section, we find that our simulated galaxy population reproduces many of the features of observed galaxy correlation functions: a power-law form with  $\gamma \approx 1.75$  down to separations  $r \sim 0.02 h^{-1}$  Mpc, an increase of correlation strength with mass that becomes more pronounced at high masses, and a higher correlation amplitude for galaxies with older stellar populations. The trends of clustering strength with galaxy mass and age agree with the predictions from semianalytic models (Kauffman et al. 1999a; Benson et al. 2000a; Somerville et al. 2001) and other hydrodynamic simulations (Cen & Ostriker 2000; Pearce et al. 2001; Yoshikawa et al. 2001). The predicted correlation lengths are somewhat low and the  $\xi(r)$  slope for  $L_*$  galaxies slightly too steep in comparison with the SDSS results, but our  $N$ -body experiments suggest that these discrepancies are largely explained by the truncation of power at the scale of our  $50 h^{-1}$  Mpc box and by a lower than average clustering amplitude in the particular  $\Lambda$ CDM initial conditions used in our simulation.

### 3.3. Pairwise Velocities

Peebles (1980), Davis & Peebles (1983), and Bean et al. (1983) describe the techniques now commonly used in correlation function analyses of galaxy redshift surveys. Starting from  $\xi(r_p, \pi)$ , the correlation as a function of projected and line-of-sight separations, one integrates over  $\pi$  to obtain the projected correlation function  $w(r_p)$ , which can be easily inverted to obtain the real-space three-dimensional correlation function  $\xi(r)$ . The full  $\xi(r_p, \pi)$  can be calculated, approximately, by a convolution of  $\xi(r)$  with the galaxy pairwise velocity distribution (see Fisher 1995 for a careful discussion of this approximation and its relation to the linear theory approximation of Kaiser 1987 and Hamilton 1992). This distribution is specified by its first and second moments,  $v_{12}(r)$  and  $\sigma_{12}^2(r)$ , and by a functional form, usually assumed (on both theoretical and observational grounds) to be exponential. The pairwise dispersion can be inferred from  $\xi(r_p, \pi)$  in a

fairly robust fashion, although it is sensitive to the presence of rare, rich clusters in the survey volume (Mo, Jing, & Börner 1993). The mean pairwise velocity can in principle be deduced as well, given sufficiently good data (see, e.g., the analysis of Fisher et al. 1994); at larger separations, it can also be measured directly in galaxy peculiar velocity surveys (Ferreira et al. 1999).

Figure 8 shows the mean pairwise velocity (*top*) and pairwise velocity dispersion (*bottom*) at  $z = 0$  (*left*) and  $z = 1$  (*right*). At  $z = 1$ , the mean pairwise velocity of the galaxies is higher than that of the dark matter, at all separations  $0.1\text{--}10 h^{-1}$  Mpc. Although galaxies and dark matter feel the same gravitational forces and therefore have similar large-scale velocity fields, pair-weighted statistics such as  $v_{12}(r)$  and  $\sigma_{12}(r)$  are also sensitive to spatial bias; in this case the preferential location of galaxies in overdense regions with high inflow velocities boosts  $v_{12}(r)$ . The mean pairwise velocities of the  $L_*$  sample are higher than those of the complete sample, but the difference is of marginal statistical significance.

By  $z = 0$  (*top left*), the bias of the galaxy mean pairwise velocities has largely disappeared, corresponding to the small spatial bias at  $z = 0$  (see Figs. 4 and 9). The complete and  $L_*$  galaxy samples again exhibit similar  $v_{12}(r)$ , with velocities of the  $L_*$  sample being slightly higher. There is a much sharper difference, however, between the older (Fig. 8, *circles*) and younger (Fig. 8, *squares*) halves of the complete sample, with the older galaxies having nearly twice the value of  $v_{12}(r)$  on scales of one to several  $h^{-1}$  Mpc. This difference reflects the preference of older galaxies for denser environments with larger inflow velocities.

The pairwise dispersion of the complete galaxy sample is below that of the dark matter at both  $z = 1$  and 0. The  $L_*$  sample's pairwise dispersion is similar to that of the dark matter at  $z = 1$  and below it at  $z = 0$ . All the galaxy samples show a peak in  $\sigma_{12}(r)$  at  $r \sim 0.3 h^{-1}$  Mpc, roughly the half-mass radius of the richest groups. While one might expect suppression of the galaxy pairwise dispersion to go hand in hand with spatial antibias, our results show that the story is not so simple. For example, the correlation function of the  $L_*$  sample at  $z = 0$  is positively biased at all separations (Fig. 6), but its pairwise dispersion is lower than that of the dark matter and nearly identical to that of the (more weakly clustered) complete galaxy sample. As emphasized by Berlind & Weinberg (2002) and Sheth et al. (2001), the pairwise dispersion is influenced by factors that do not affect  $\xi(r)$  (or even affect it with opposite sign), making it sensitive to details of the relation between galaxies and their surrounding dark matter halos. While  $\sigma_{12}(r)$  is not sensitive to galaxy baryon mass (at least in the comparison presented here), it is highly sensitive to stellar

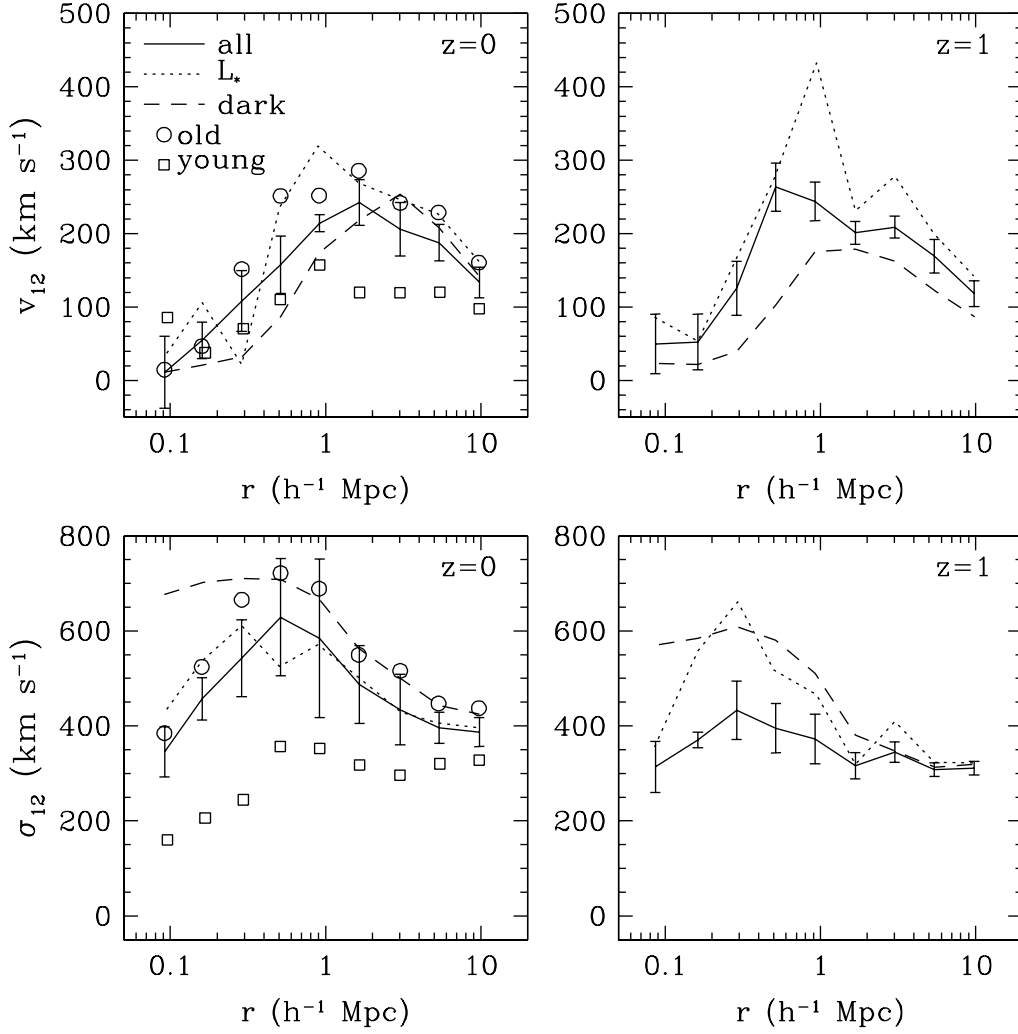


FIG. 8.—Mean pairwise radial velocities (*top*) and pairwise radial velocity dispersions (*bottom*) at  $z = 0$  and  $z = 1$ . Solid lines with jackknife error bars show results for the complete galaxy sample, while dotted and dashed lines show results for the  $L_*$  sample and the dark matter, respectively. The  $z = 0$  panels show results for the older (*circles*) and younger (*squares*) halves of the complete sample. The pairwise velocity is defined with Hubble flow removed and uses the convention of positive sign for galaxies moving toward each other. The pairwise dispersion is one-dimensional, along the direction of separation, with the mean pairwise velocity subtracted.

population age: the older half of the complete sample has a substantially higher pairwise dispersion than the younger half because of the location of older galaxies in more massive virialized structures.

The pairwise dispersion of the complete and  $L_*$  galaxy samples agrees reasonably well with the values  $\sigma_{12} \sim 550\text{--}650$  km s<sup>-1</sup> at  $r \sim 1$  h<sup>-1</sup> Mpc found in the CfA2, LCRS, and SDSS redshift surveys (Marzke et al. 1995; Jing et al. 1998; Zehavi et al. 2002). However, the sensitivity of  $\sigma_{12}(r)$  to rare, rich clusters is reflected in the large jackknife error bars of Figure 8, so we cannot take this agreement as a major triumph of the simulation. The scale dependence in Figure 8 contrasts with the rather flat  $\sigma_{12}(r)$  inferred from the observations, but for the same reason we do not view this as a serious discrepancy. We can conclude from Figure 8 that the predicted difference between dark matter and galaxy pairwise dispersions makes it easier to understand the relatively low observed value of  $\sigma_{12}$ , and that the predicted difference between old and young galaxies naturally accounts for the factor of  $\sim 2$  difference in  $\sigma_{12}$  derived for red and blue galaxies in the SDSS (Zehavi et al. 2002) and the factor of  $\sim 1.7$  difference in  $\sigma_{12}$  derived from

optical redshift surveys (as cited above) and from the IRAS PSCz catalog (as analyzed by Jing, Börner, & Suto 2002).

Our conclusion that the galaxy pairwise dispersion at  $z = 0$  is significantly lower than that of the dark matter agrees with the results of Pearce et al. (2001), which are based on a hydrodynamic simulation, and Benson et al. (2000a), which are based on populating  $N$ -body halos according to semianalytic prescriptions. We also agree fairly well with these calculations in terms of the amplitude of the pairwise dispersion,  $\sigma_{12} \sim 500\text{--}600$  km s<sup>-1</sup> at  $r \sim 1$  h<sup>-1</sup> Mpc, and the shape of  $\sigma_{12}(r)$ , although our pairwise dispersion drops somewhat more rapidly at  $r < 0.5$  h<sup>-1</sup> Mpc. Kauffmann et al. (1999a), also using  $N$ -body halos and semianalytic galaxy models, find only a slight difference between the dark matter and galaxy pairwise dispersions, and their predicted amplitude at  $r \sim 1$  h<sup>-1</sup> Mpc is  $\sigma_{12} \sim 800$  km s<sup>-1</sup>. Benson et al. (2000a) argue that the difference between their results and those of Kauffmann et al. arises because their (Benson et al.) semianalytic model places fewer galaxies in high dispersion halos. Since the halo occupation of our simulation agrees well with that of the Benson et al. (2000a) models (Berlind et al. 2003), we think that the

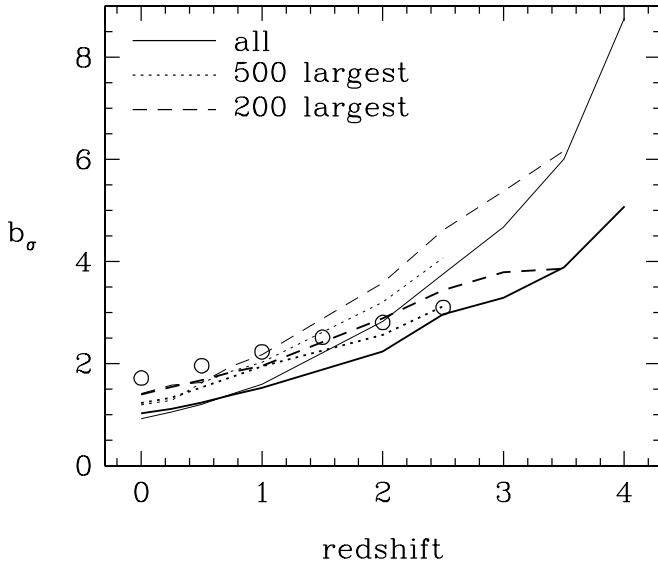


FIG. 9.—Evolution of the bias factor  $b_\sigma$ , defined by the ratio of rms galaxy count fluctuations (corrected for shot noise) to rms dark matter fluctuations in spheres of comoving radius  $8 h^{-1}$  Mpc (heavy lines) or  $2 h^{-1}$  Mpc (light lines), showing results for the complete galaxy sample (solid lines), the 500 most massive galaxies (dotted lines), and the 200 most massive galaxies (dashed lines). Lines for the 500 largest galaxies stop at  $z = 2.5$ , since the complete sample contains fewer than 500 galaxies at higher redshift; likewise, lines for the 200 largest galaxies stop at  $z = 3.5$ . Circles show the number-conserving model of Fry (1996), normalized to a bias factor of 3.1 at redshift  $z = 2.5$ .

same explanation applies to our results. We agree with Kauffmann et al. (1999a) with respect to the mean pairwise velocity (not examined by the other groups), with both calculations showing little difference between the galaxies and the dark matter at  $z = 0$ . While age divisions (or color divisions) were not examined in these other studies, the tendency to find older, redder galaxies in more massive halos appears to be a generic prediction of both hydrodynamic and semianalytic calculations, so we expect the strong age dependence of  $\sigma_{12}(r)$  and  $v_{12}(r)$  to be a fairly generic result.

### 3.4. Moments of Galaxy Counts

Moments of the galaxy count distribution in spherical (or other) cells provide an alternative to correlation functions for characterizing galaxy clustering. The  $n$ th central moment of the galaxy counts is closely related to the  $n$ -point correlation function averaged over the cell volume. As pointed out by Szapudi & Szalay (1993), for a discrete galaxy distribution it is convenient to work with factorial moments such as  $\langle N(N-1) \rangle$  and  $\langle N(N-1)(N-2) \rangle$  (where  $N$  is the number of galaxies within a cell or sphere), since these automatically remove Poisson shot noise contributions and can therefore be related to the moments of an underlying continuum distribution in a straightforward way.

The ratio  $b_\sigma$  of rms galaxy count fluctuations to rms mass fluctuations provides a simple and robust measure of galaxy bias, and the value of  $b_\sigma$  for spheres of radius  $8 h^{-1}$  Mpc is the number most often quoted as the “bias factor” of the galaxy distribution. Figure 9 shows  $b_\sigma(8 h^{-1} \text{ Mpc})$  and  $b_\sigma(2 h^{-1} \text{ Mpc})$  as a function of redshift, with

$$b_\sigma \equiv \left[ \frac{\langle N(N-1) \rangle}{\langle N \rangle^2} - 1 \right]^{1/2} \left\langle \left( \frac{\delta\rho}{\rho} \right)^2 \right\rangle^{-1/2} \quad (3)$$

and the moments evaluated using  $10^6$  spheres placed randomly in the simulation volume. At  $z = 0$  the bias factors are similar for  $r = 2$  and  $8 h^{-1}$  Mpc, with  $b_\sigma(8 h^{-1} \text{ Mpc}) = 1.02$  for the complete galaxy sample, 1.22 for the  $L_*$  sample, and 1.40 for the 200 most massive galaxies. Figures 5–8 show that the simulated galaxies do not trace the mass in detail and that galaxy clustering depends on age and luminosity, but our complete galaxy sample is, in an rms sense, nearly unbiased at  $z = 0$ .

As shown earlier in terms of  $\xi(r)$ , the galaxy bias starts high and declines with redshift as the dark matter clustering catches up. For the complete sample,  $b_\sigma(8 h^{-1} \text{ Mpc}) = 3.3$  at  $z = 3$ . At redshifts  $z \sim 1$ , the bias factor is significantly larger at  $2 h^{-1}$  Mpc than at  $8 h^{-1}$  Mpc, presumably reflecting the different behavior of  $b_\sigma$  in the nonlinear and linear regimes. Fry (1996) discusses an analytic model in which the galaxy population is born at some redshift  $z_*$  with a bias factor  $b_*$  but thereafter moves with the same velocity as the dark matter and conserves number (see also Tegmark & Peebles 1998). In this case the bias factor at redshift  $z < z_*$  is given by  $b(z) - 1 = (b_* - 1)D(z_*)/D(z)$ , where  $D(z)$  is the linear growth factor at redshift  $z$ . Circles in Figure 9 show the prediction of this model, normalized to the bias factor  $b_* = 3.1$  of the 500 galaxy sample at  $z_* = 2.5$ . The actual bias factor of this sample declines significantly faster than the number-conserving model predicts, even though the sample contains 500 galaxies at all redshifts, presumably because mergers drive down the number of pairs in dense regions and galaxies that form at  $z < 2.5$  are born in less biased environments. Somerville et al. (2001) obtain a similar result with semianalytic modeling.

Discussions of the third moment of the count distribution often focus on the hierarchical ratio  $S_3 \equiv \langle \delta^3 \rangle / \langle \delta^2 \rangle^2$ . For spherical cells and a linear power spectrum  $P(k) \propto k^n$ , the dark matter has  $S_3 = 34/7 - (3+n)$  to second order in perturbation theory, independent of sphere radius and power spectrum amplitude and nearly independent of  $\Omega_m$  (Juszkiewicz, Bouchet, & Colombi 1993). If the galaxy density contrast is a local function of the matter density contrast,  $\delta_g = f(\delta)$ , then the relation between  $S_{3g}$  and  $S_3$  (representing the galaxy and mass moments, respectively) depends on both the linear bias factor  $b_\sigma$  and a second bias factor  $b_2$  that characterizes the second derivative  $f''(\delta)$  at  $\delta = 0$  (Fry & Gaztañaga 1993; Juszkiewicz et al. 1995). The behavior of  $S_3$  and the relation between  $S_{3g}$  and  $S_3$  becomes more complicated in the fully nonlinear regime in which second-order perturbation theory breaks down, but it remains the case that  $S_3$  is only weakly dependent on scale and that the ratio  $S_{3g}/S_3$  depends on the nonlinearity of the relation between galaxy and mass densities.

Figure 10 shows the hierarchical ratio  $S_{3f}$  defined in terms of factorial moments,

$$S_{3f} \equiv \left[ \frac{\langle N(N-1)(N-2) \rangle}{\langle N \rangle^3} - 1 \right] \left[ \frac{\langle N(N-1) \rangle}{\langle N \rangle^2} - 1 \right]^{-2}, \quad (4)$$

at  $z = 0$ . For the complete galaxy sample,  $S_{3f} \approx 3.5$ – $4$  for  $r = 1$ – $4 h^{-1}$  Mpc, rising to  $S_{3f} = 6.5$  at  $r = 8 h^{-1}$  Mpc. These values are slightly below those of the dark matter distribution. Ratios for the more massive,  $L_*$  sample are lower still; these galaxies are more clustered in an rms sense, but they are less “skewed” in the sense measured by  $S_{3f}$ . As emphasized by Colombi et al. (2000), the statistical and systematic errors in measurements of  $S_3$  from a volume this size are substantial,

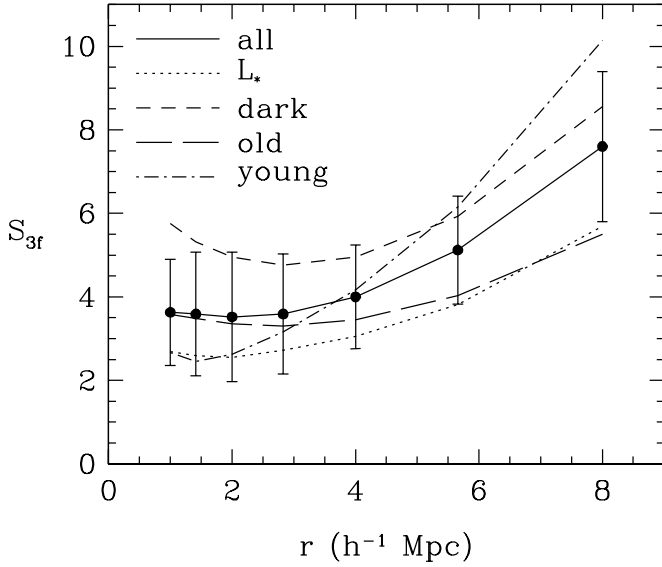


FIG. 10.—Hierarchical amplitude  $S_{3f}$ , defined as the ratio of the third factorial moment to the square of the second factorial moment in spheres of radius  $r$ , at  $z = 0$ . Points connected by the solid line show results for the complete galaxy sample, with jackknife error bars. Other lines show results for the dark matter (short-dashed line), the  $L_*$  sample (dotted line), and the older (long-dashed line) and younger (dot-dashed line) halves of the complete sample. The value of  $S_3$  inferred from angular clustering in the APM galaxy catalog is roughly constant at  $S_3 \approx 3.5$  over this range of scales.

and our jackknife error bars are large. Thus, while the predicted value of  $S_{3f}$  for the complete galaxy sample agrees well with the value  $S_3 \approx 3.5$  that Gaztañaga & Frieman (1994) infer by deprojecting the angular count moments in the APM galaxy catalog, we cannot claim this as a major success. (The recent analyses of angular clustering in the SDSS by Gaztañaga 2002 and Szapudi et al. 2002 suggest that the APM numbers should be adjusted upward by  $\sim 20\%$ .) However, the relative values of  $S_{3f}$  for the different populations should be more robustly predicted, and our simulation indicates that galaxies comparable to those in the complete sample should have a hierarchical ratio slightly below that of the dark matter and that  $S_{3f}$  should be smaller for more massive galaxies. The most striking result in Figure 10 is the strong scale dependence of  $S_{3f}$  for young galaxies, a prediction that should be testable with the SDSS and 2dFGRS surveys in the near future.

#### 4. GALAXY-MASS CORRELATIONS

Advances in wide-field CCD imaging have opened a new window on the relation between galaxies and mass: weak-lensing measurements of the galaxy-mass cross-correlation function. This statistic can be derived from galaxy-galaxy lensing analyses, which measure the mean shear profile around a sample of foreground galaxies by using the tangential distortion of background galaxies (Tyson et al. 1984; Brainerd, Blandford, & Smail 1996; dell’Antonio & Tyson 1996; Griffiths et al. 1996; Hudson et al. 1998; Fischer et al. 2000; Hoekstra, Yee, & Gladders 2001; McKay et al. 2001; Smith et al. 2001; Wilson et al. 2001b). Alternatively, one can correlate the mass map derived from the ellipticity correlation function of faint galaxies with the light map of brighter, foreground galaxies in the same images (Wilson, Kaiser, & Luppino 2001a). In addition, cosmic shear measurements are beginning to yield measurements of the matter correlation function

itself (for specified  $\Omega_m$ ), with impressively good agreement among independent surveys (see Mellier et al. 2001; Hoekstra et al. 2002 and references therein), allowing for the comparison of galaxy and mass fluctuation amplitudes.

Figure 11 shows the galaxy-mass cross correlation function for the complete galaxy sample and for the subsamples divided by age (*top*), mass (*middle*), and local density (*bottom*). We define  $\xi_{gm}(r) = \langle \delta_g(\mathbf{x})\delta_m(\mathbf{x} + \mathbf{r}) \rangle$ , where  $r = |\mathbf{r}|$  and  $\delta_g = n_g/\bar{n}_g - 1$  and  $\delta = \rho/\bar{\rho} - 1$  are the galaxy and mass (dark+baryonic) overdensities; we compute it by counting galaxy-particle pairs, with particles weighted by mass. While the form of  $\xi_{gm}(r)$  at small  $r$  is connected to the typical galaxy halo profile, the connection is rather loose because the

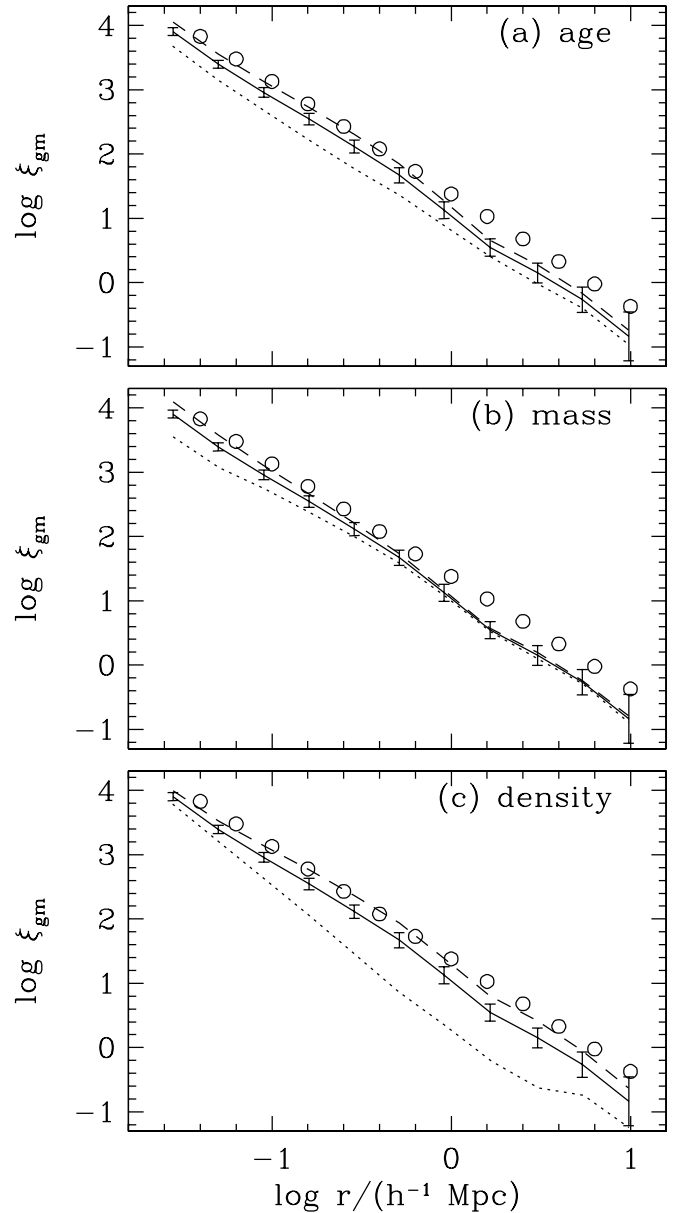


FIG. 11.—Cross-correlation function  $\xi_{gm}(r)$  of galaxies and mass (dark matter+baryons) at  $z = 0$ . In each panel, the solid line with jackknife error bars shows results for the complete galaxy sample. Dashed lines show results for the older (a), more massive (b), and higher density half of the complete sample (c). Dotted lines show results for the other half of the sample in each panel. Circles are the same as those in Fig. 7, to facilitate visual comparison with the autocorrelation functions.

galaxies are not isolated and the samples contain objects with a range of halo virial radii (points emphasized by Guzik & Seljak 2001, White et al. 2001, and Yang et al. 2003). We use the same format in Figures 11 and 7, and the two figures bear a striking resemblance to each other. To a first approximation, the galaxy-mass correlation function has the same shape as the galaxy-galaxy correlation function, similar amplitude, and similar dependence on galaxy age, baryon mass, and environment. However, the variations of  $\xi_{gm}(r)$  for different subsamples are smaller than the variations of  $\xi_{gg}(r)$ , as one would expect, given that the mass distribution stays the same regardless of the galaxy subsample.

With measurements of  $\xi_{gg}(r)$ ,  $\xi_{gm}(r)$ , and the mass correlation function  $\xi_{mm}(r)$ , one can define two “bias functions,”  $b_\xi(r) \equiv [\xi_{gg}(r)/\xi_{mm}(r)]^{1/2}$  and  $b_{\xi\times}(r) \equiv \xi_{gg}(r)/\xi_{gm}(r)$ . The ratio  $r_\times(r) \equiv b_\xi(r)/b_{\xi\times}(r) = \xi_{gm}(r)/[\xi_{gg}(r)\xi_{mm}(r)]^{1/2}$  corresponds to a “correlation coefficient” as defined, e.g., by Tegmark & Bromley (1999), although here we work in terms of correlation functions rather than power spectra. In the case of linear, deterministic bias, where  $\delta_g(\mathbf{x}) = b\delta(\mathbf{x})$  at all positions  $\mathbf{x}$ ,  $b_\xi$  and  $b_{\xi\times}$  would be equal to  $b$  and independent of scale, and  $r_\times$  would equal unity. Figure 12 plots  $b_\xi(r)$ ,  $b_{\xi\times}(r)$ , and  $r_\times(r)$  for the complete galaxy sample, the  $L_*$  sample, and the older and the younger halves of the complete sample. We focus on the  $z = 0$  results, but because measurements of these quantities come from deep imaging surveys, we also show the simulation predictions at  $z = 0.5$ , which are qualitatively similar.

The complete sample has  $b_\xi \approx 1$ , but there is a broad minimum around  $0.5 h^{-1}$  Mpc, with  $b_\xi(0.5 h^{-1} \text{ Mpc}) \approx 0.75b_\xi(10 h^{-1} \text{ Mpc})$ . It is precisely this mild scale dependence of  $b_\xi(r)$  that transforms the curved dark matter correlation function into a power-law galaxy correlation function (see Fig. 6). The  $L_*$  sample has higher  $b_\xi$ , roughly flat (though noisy) at  $r \gtrsim 0.5 h^{-1}$  Mpc, climbing toward small  $r$ . The old galaxies have  $b_\xi(r)$  similar to that of the  $L_*$  sample, while the young galaxies have low  $b_\xi(r)$  with a scale dependence similar to that of the complete sample.

The scale dependence of  $b_{\xi\times}(r)$  is generally weaker than that of  $b_\xi(r)$ . At  $r \gtrsim 1 h^{-1}$  Mpc, values of  $b_{\xi\times}$  are similar to those of  $b_\xi$  for the corresponding sample, except in the case of the young galaxies, which have  $b_{\xi\times} > b_\xi$  and thus  $r_\times(r) < 1$ . This “anticorrelation” arises because the young galaxies tend to avoid the higher mass halos, which contribute significantly to  $\xi_{mm}(r)$ . The complete,  $L_*$ , and old galaxy samples have  $r_\times(r) \approx 1$  at large scales and  $r_\times(r) > 1$  at small scales. While a correlation coefficient defined in terms of smoothed density fields, e.g.,  $\langle \delta_g(\mathbf{x})\delta_m(\mathbf{x}) \rangle / [\langle \delta_g^2(\mathbf{x}) \rangle \langle \delta_m^2(\mathbf{x}) \rangle]^{1/2}$ , must lie in the range  $[-1, 1]$ , the same is not true for  $r_\times(r)$ , which is defined by ratios of correlation functions. The high values of  $r_\times(r)$  at small  $r$  reflect the location of a significant fraction of galaxies (especially in the  $L_*$  sample) near the central density maxima of dark matter halos. Galaxy-mass and galaxy-galaxy pairs involving these central galaxies boost  $\xi_{gm}(r)$  and  $\xi_{gg}(r)$ , respectively, relative to the values expected if galaxies traced mass within halos. There is no corresponding boost to  $\xi_{mm}(r)$ , and since  $r_\times(r) \propto \xi_{gm}(r)/[\xi_{gg}(r)]^{1/2}$ , it tends to rise above unity on scales where central galaxies make major contributions to the correlation functions (for related discussions, see Seljak 2000 and Berlind & Weinberg 2002).

Hoekstra et al. (2001) measure a ratio of galaxy-mass and galaxy-galaxy fluctuations that is directly analogous to  $b_{\xi\times}(r)$ , although it is defined in compensated apertures rather than annuli  $r \rightarrow r + \Delta r$ . In our terminology, their results imply

$b_{\xi\times} = 1.26_{-0.12}^{+0.15}$  for our flat,  $\Omega_m = 0.4$  cosmology (converted from their  $b/r_\times = 1.05$  for  $\Omega_m = 0.3$  by using their eq. [14]), with no detectable scale dependence in the range  $r \sim 0.1$ – $1.5 h^{-1}$  Mpc. This result is intermediate between that of our complete and  $L_*$  galaxy samples at  $z = 0.5$ . (The median redshift of the Hoekstra et al. lens sample is  $z \sim 0.35$ , where the bias should be intermediate between our  $z = 0$  and  $0.5$  values.) While we have not modeled the Hoekstra et al. procedure in detail, our prediction of a mild positive bias with no substantial scale dependence and a bias factor similar to that applying on larger scales appears to be fully compatible with their results. The approximate scale independence of  $b_{\xi\times}(r)$  also agrees with the finding of Wilson et al. (2001a) of similar shapes for the galaxy-galaxy and galaxy-mass angular correlation functions. It is difficult for us to compare the amplitude of  $b_{\xi\times}$  with the result of Wilson et al., in part because they weight galaxies by flux and in part because they interpret their measurement in terms of a value of  $\Omega_m$  assuming that their (red luminous) lens galaxy population traces the mass. The existence of positively and negatively biased samples with approximately scale-independent  $b_{\xi\times}(r)$  in Figure 11 demonstrates that a constant ratio  $\xi_{gg}(r)/\xi_{gm}(r)$  does not, on its own, justify this assumption.

One can also use galaxy-galaxy lensing to infer the average mass within a sphere of radius  $r$  centered on a sample galaxy, an approach emphasized by McKay et al. (2001). (Note that this measurement makes no attempt to distinguish mass that “belongs” to the central galaxy from mass that “belongs” to its neighbors or to the common halo of its group or cluster; all mass counts, and it may be counted multiple times for different galaxies.) Figure 13a shows the average mass profile from  $25 h^{-1}$  kpc to  $1 h^{-1}$  Mpc in five different bins of galaxy baryon mass, each a factor of 2 in width. We express the mass in terms of an equivalent circular velocity  $v_c = [G\Delta M(r)/r]^{1/2}$ , which would be constant in the case of a singular isothermal sphere, although of course few galaxies would have any tracers in circular rotation at these large radii. We use the excess mass  $\Delta M$ , subtracting the mean value  $4\pi r^3 \bar{\rho}/3$ , because it is the excess mass that is constrained by lensing, but the difference is negligible on the scale of these plots.

For the lowest mass galaxies,  $M_b = (0.5 - 1) \times 10^{11} M_\odot$ , the mean circular velocity curve is fairly flat at  $v_c \sim 200 \text{ km s}^{-1}$  out to  $r \sim 100 h^{-1}$  kpc, but it rises at larger radii because these galaxies typically reside in groups whose characteristic circular velocities are larger than those of the galaxies themselves. Conversely, the highest mass galaxies have a mean circular velocity curve that falls all the way out to  $r = 1 h^{-1}$  Mpc, because these massive galaxies typically reside at the centers of potential wells whose mass profiles (baryon+dark) are steeper than isothermal. Intermediate mass ranges have intermediate behavior, with falling  $v_c$  profiles at small  $r$ , changing to flat or rising profiles at  $r \sim 100$ – $200 h^{-1}$  kpc. Dashed and dotted curves show  $v_c(r)$  for the older and younger halves, respectively, of the  $(1 - 2) \times 10^{11} M_\odot$  sample. At small radii ( $r \lesssim 50 h^{-1}$  kpc), the circular velocity is the same for the two subsamples, but the preferential location of older galaxies in more massive groups manifests itself as a higher  $v_c$  at larger radii.

Figure 13b plots the ratio of the aperture mass  $\Delta M(r)$  to galaxy baryon mass  $M_b$  (which is approximately equal to the stellar mass) as a function of  $M_b$ . From bottom to top, solid points connected by solid lines show results for  $r = 50, 100, 260, 500,$  and  $1000 h^{-1}$  kpc. We use  $260 h^{-1}$  kpc rather than  $250 h^{-1}$  to facilitate comparison with the analysis of galaxy-

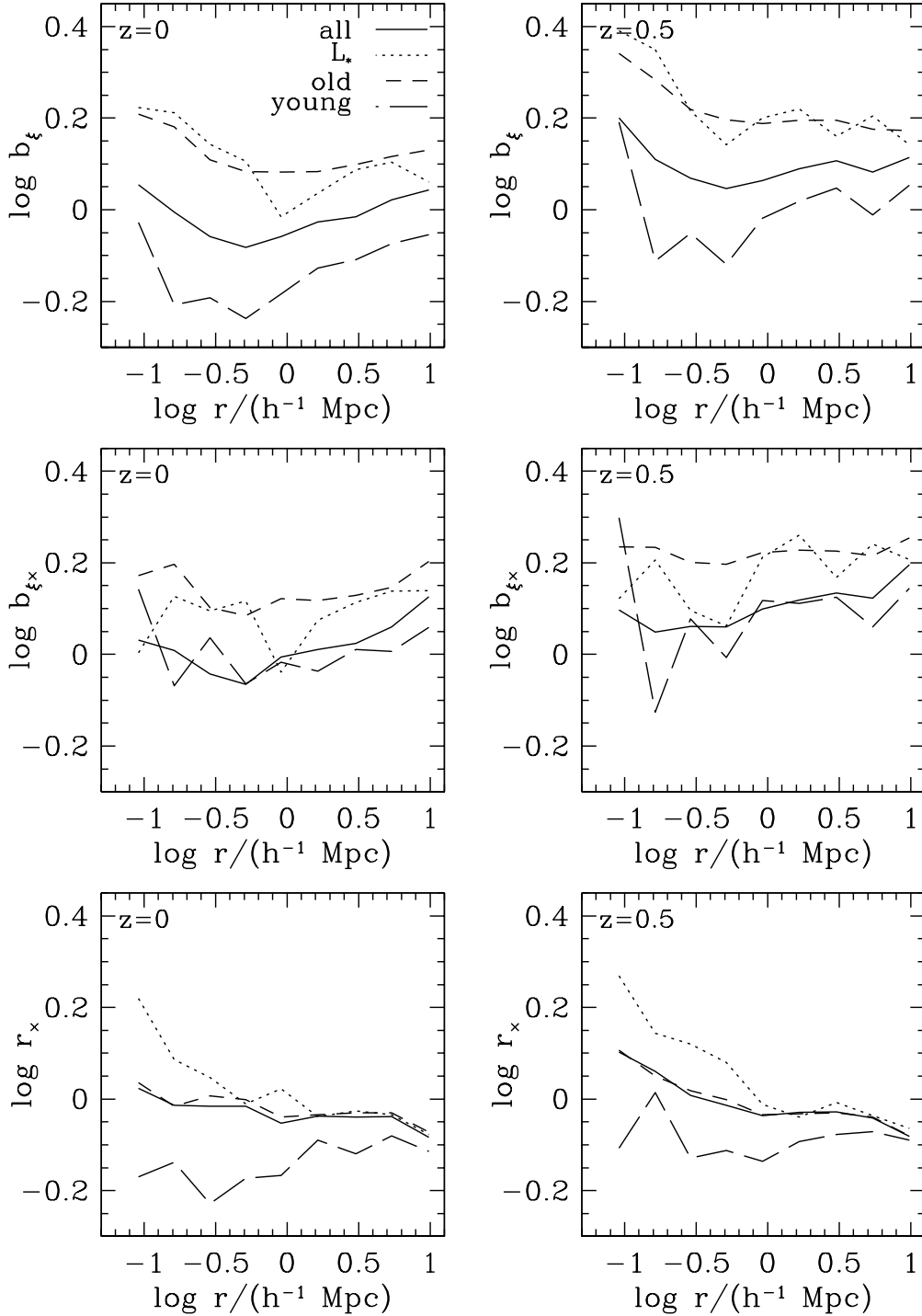


FIG. 12.—Scale dependence of the bias factors defined by  $b_\zeta(r) = [\xi_{gg}(r)/\xi_{mm}(r)]^{1/2}$  (top) and by  $b_{\zeta_x}(r) = \xi_{gg}(r)/\xi_{gm}(r)$  (middle), and the “correlation coefficient”  $r_x(r) = b_\zeta(r)/b_{\zeta_x}(r)$  (bottom), where  $\xi_{gg}(r)$  and  $\xi_{mm}(r)$  are the galaxy and mass autocorrelation functions and  $\xi_{gm}(r)$  is the galaxy-mass cross-correlation function. Results are shown for the complete galaxy sample (solid lines), the  $L_*$  sample (dotted line), the older (short-dashed line) and younger (long-dashed line) halves of the complete sample, at redshifts  $z = 0$  (left column) and  $z = 0.5$  (right column).

galaxy lensing in the SDSS of McKay et al. (2001). We average in the order appropriate for such a study, computing the mean value of the excess mass in bins of  $M_b$ , dividing by the mean  $M_b$  of the bin, and plotting the result for each bin at the value of the mean  $M_b$ . Note that this ratio counts all excess mass within radius  $r$ , including that associated with neighboring galaxies, but divides by the baryon mass of the central galaxy only. For  $r = 50 h^{-1}$  kpc the  $\Delta M(r)/M_b$  curve is essentially flat—on average, a galaxy’s baryon mass is

proportional to the total mass in a  $50 h^{-1}$  kpc sphere around it. For larger  $r$ , the ratio rises toward smaller  $M_b$ , which makes sense if the apertures around these galaxies start to include more mass that “belongs” to their richer neighbors. Dotted and dashed lines are computed, respectively, for the older and younger halves of the galaxies in each  $M_b$  bin, at  $r = 260 h^{-1}$  kpc. For  $M_b \lesssim 5 \times 10^{11} M_\odot$ , the older galaxies have substantially higher  $\Delta M(260 h^{-1} \text{ kpc})$ , by a factor  $\sim 1.5$ – $2$ , analogous to the factor  $\sim 2.7$  difference that McKay et al. (2001)



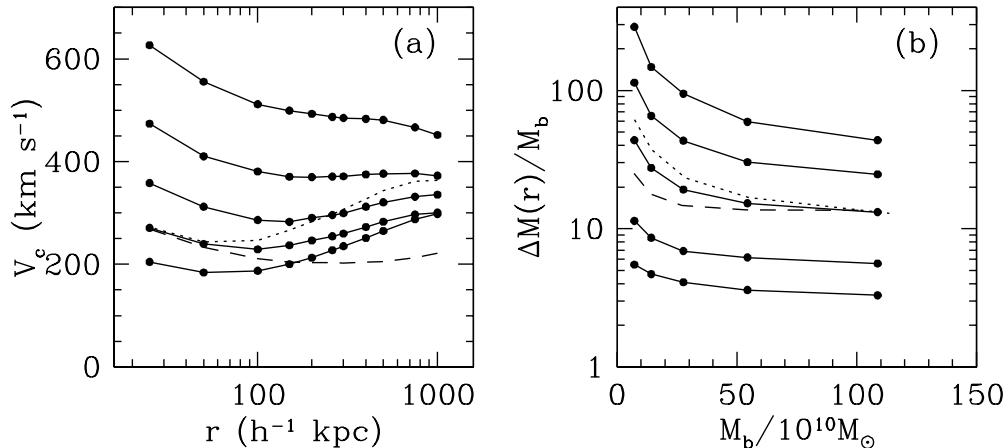


FIG. 13.—Mass distribution around galaxies at  $z = 0$ . (a) Circular velocity  $v_c(r) \equiv [G\Delta M(r)/r]^{1/2}$  as a function of radius where  $\Delta M(r)$  is the mean total excess mass in spheres of radius  $r$  centered on galaxies (solid curves). Solid curves, bottom to top, Results for galaxies with baryon mass (stars plus cold gas) in the range 0.5–1, 1–2, 2–4, 4–8, and 8–16, in units of  $10^{11} M_\odot$ ; results for older (dotted curve) and younger (dashed curves) halves of the galaxies in the  $(1-2) \times 10^{11} M_\odot$  mass range. (b) Ratio of the mean total excess mass within radius  $r$  to the baryon mass of the central galaxy as a function of the galaxy baryon mass, for radii  $r = 50, 100, 260, 500,$  and  $1000 h^{-1} \text{kpc}$  (solid curves, bottom to top); results at  $r = 260 h^{-1} \text{kpc}$  for galaxies that are older (dotted curve) and younger (dashed curve) than the median age of galaxies in the mass bin.

find for elliptical and spiral subsamples. At higher masses, there is little difference in age between the older and younger halves of the sample (essentially all massive galaxies are old), and the systematic difference in  $\Delta M(260 h^{-1} \text{kpc})$  narrows.

Figure 14 compares our predictions with the results of McKay et al. (2001), who estimate  $\Delta M(260 h^{-1} \text{kpc})$  by fitting isothermal spheres to their measured shear profiles in four wide bins of galaxy luminosity. They find that  $\Delta M(260 h^{-1} \text{kpc})/L$  is roughly independent of galaxy luminosity  $L$  in the Sloan  $g$ -,  $r$ -,  $i$ -, and  $z$ -bands. To convert  $\Delta M(260 h^{-1} \text{kpc})/L$  to  $\Delta M(260 h^{-1} \text{kpc})/M_b$  for comparison with our predictions, we draw on the results of Kauffman et al. (2003), who estimate ratios of stellar mass to observed luminosity by fitting population synthesis and dust extinction models to SDSS galaxy spectra and colors. We compute the mean  $M_{\text{stel}}/L$  in bins of absolute magnitude by using tabulated  $M_{\text{stel}}/L$  distributions kindly provided by G. Kauffmann, excluding the 10% tails of the distributions so that the mean is not distorted by extreme  $M_{\text{stel}}/L$  values that may be a result of poor model fits. We take the values of  $\Delta M(260 h^{-1} \text{kpc})/L$  from Table 2 of McKay et al., convert to  $h = 0.65$ , and divide by the value of  $\langle M_{\text{stel}}/L \rangle$  at the central luminosity of each bin (which McKay et al. compute using the same galaxy weights that enter the lensing analysis). In Figure 14, triangles, squares, pentagons, and circles show results derived from the  $g$ -,  $r$ -,  $i$ -, and  $z$ -band mass-to-light ratios, respectively.<sup>4</sup> Results in the four bands are not independent because they rely on the same lensing data, but the generally good agreement among them suggests that there are no major systematic errors in the population synthesis and dust extinction modeling. The data points come in four groups reflecting the four luminosity bins, although the  $g$ -band points do not reach the highest  $M_b$  values because of the higher  $M_{\text{stel}}/L$  in  $g$  band. The solid line in Figure 14 shows the simulation result for all galaxies (same as Fig. 13b, middle solid line).

<sup>4</sup> This plot is noticeably different from the one in the preprint version of this paper because of revisions in the estimated  $M_{\text{stel}}/L$  ratios between the preprint and published versions of Kauffman et al. (2003). These revisions, which arose from the use of improved spectral synthesis models, had little impact at most galaxy luminosities, but they changed  $M_{\text{stel}}/L$  by as much 25%–50% at the high luminosities in the top two McKay et al. (2001) bins.

The predicted and observationally inferred values of  $\Delta M(260 h^{-1} \text{kpc})/M_b$  agree well in the lowest mass bin, but at higher masses the observational points lie above the simulation predictions, by a factor  $\sim 1.5$ – $3$  on average. We have treated our galaxies as entirely stellar, and we have ignored the light “missed” by SDSS Petrosian luminosities (see Strauss et al. 2002), but accounting for these effects would reduce the discrepancy by only  $\sim 10\%$ – $20\%$ . We conclude that the baryon masses of our simulated galaxies are on average too high by a

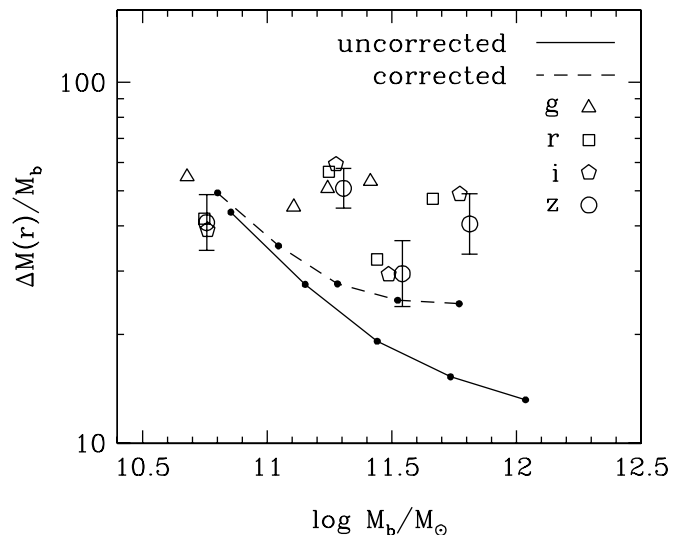


FIG. 14.—Comparison of the simulation results for  $\Delta M(260 h^{-1} \text{kpc})/M_b$  with observational estimates from the SDSS. The solid curve (same as the central solid curve in Fig. 13b) shows the prediction from our  $(50 h^{-1} \text{Mpc})^3$  simulation. The dashed curve shows this prediction with baryon masses rescaled to account for numerical resolution effects, on the basis of our comparison of higher resolution simulations of smaller volumes. It represents our best guess at the result we would obtain from a simulation using  $2 \times 288^3$  particles instead of  $2 \times 144^3$ . Points show the weak-lensing results of McKay et al. (2001), converted from  $\Delta M(260 h^{-1} \text{kpc})/L$  to  $\Delta M(260 h^{-1} \text{kpc})/M_b$  by using the estimates of  $M_{\text{stel}}/L$  as a function of  $L$  from Kauffman et al. (2003). Estimates from  $g$ -,  $r$ -,  $i$ -, and  $z$ -bands are represented by triangles, squares, pentagons, and circles, respectively. We plot error bars on the  $z$ -band points only, but they are similar for other bands.

factor  $\sim 1.5\text{--}3$ , for  $M_b \gtrsim 2 \times 10^{11} M_\odot$ , provided that the Kroupa (2002) stellar IMF assumed in the population synthesis modeling of Kauffmann et al. (2003) is indeed universal across galaxy types and evolutionary stages. Other hints of a similar discrepancy include the high overall fraction of stellar mass in the simulation (Davé et al. 2001) and the relatively high stellar mass-to-light ratios required to match the observed galaxy luminosity function (M. A. Fardal et al. 2004, in preparation).

We believe that these excessive galaxy baryon masses are largely numerical in origin. A comparison of two simulations of the same  $(22.22 h^{-1} \text{Mpc})^3$  volume, one with the same resolution as the  $(50 h^{-1} \text{Mpc})^3$  simulation analyzed here and one with a mass resolution 8 times higher, shows that the *same* galaxies are on average more massive in the lower resolution simulation, by  $\sim 10\%$  at the  $5 \times 10^{10} M_\odot$  completeness limit, rising to a factor  $\sim 2$  at masses of several times  $10^{11} M_\odot$  (M. A. Fardal et al. 2004, in preparation; one can see hints of this effect in the mass function comparison of Fig. 2). The higher baryon masses at low resolution probably arise from the overestimated cooling rate of hot gas particles in the vicinity of cold dense clumps (Pearce et al. 1999; Croft et al. 2001), although other effects could play a role. An alternative formulation of SPH that performs better in the presence of contact discontinuities between cold and hot phases indeed leads to lower galaxy baryon masses, typically by a factor  $\sim 2$  (Springel & Hernquist 2002). Empirically, we find that the relation  $M_{\text{tr}} = 5 \times 10^{10} M_\odot (M_{\text{tr}}/5.4 \times 10^{10} M_\odot)^{0.82}$  provides a fairly accurate scaling between masses in the “low resolution” and “high resolution” simulations. The dashed curve in Figure 14 shows the effect of applying this empirical scaling. It represents our best guess at the results we would obtain from a  $2 \times 288^3$  particle simulation of the same  $(50 h^{-1} \text{Mpc})^3$  volume. Comparison with still higher resolution simulations at  $z = 3$  suggests that this is sufficient to yield numerically converged baryon masses, but we have not yet been able to carry out such a test at  $z = 0$ .

The resolution-corrected predictions and the observationally inferred mass ratios agree well at  $\log(M_b/1 M_\odot) \sim 10.7$  and 11.5 but differ by a factor  $\sim 2$  at  $\log(M_b/1 M_\odot) \sim 11.25$  and 11.7. Clearly, there are still systematic uncertainties in the observational results (the weak-lensing measurements themselves and the  $M_{\text{stel}}/L$  scaling) and in the numerical convergence of the simulation predictions, so it is difficult to know just what to make of the agreement and the discrepancy. Overall, the current results seem to leave room for at most a factor  $\sim 2$  additional reduction in baryon masses by astrophysical processes such as active galactic nucleus feedback or stronger stellar feedback, for galaxies with  $M_b \gtrsim 5 \times 10^{10} M_\odot$ . We would probably predict lower values of  $\Delta M(260 h^{-1} \text{kpc})/M_b$  (and thus require some further suppression of baryon mass) if we adopted lower  $\Omega_m$  or higher  $\Omega_b$  without changing the assumed  $\sigma_8$  or  $P(k)$  shape. However, for our adopted cosmological parameters, the lensing comparison suggests that this simulation already incorporates the astrophysics needed to understand the baryon masses of luminous galaxies at the factor of 2 level. The constraints from galaxy-galaxy lensing complement those from the Tully-Fisher relation and direct estimates of the galaxy baryon mass function, which we discuss elsewhere (M. A. Fardal et al. 2004, in preparation).

## 5. SUMMARY AND DISCUSSION

For the full population of galaxies with  $M_b > 5 \times 10^{10} M_\odot$ , our simulation predicts a bias factor  $b_\sigma(8 h^{-1} \text{Mpc})$  of unity at

$z = 0$ . Beneath this rms similarity of fluctuation amplitudes, however, lie numerous differences between the galaxy and dark matter distributions, in terms of present-day structure and the evolution of that structure.

1. The bias defined by rms fluctuations or by correlation functions is strongly redshift dependent. If we choose a galaxy population of fixed comoving space density (above a redshift-dependent baryon mass threshold), then the galaxy correlation function stays approximately constant from  $z = 3$  to 0, so that bias is initially large and declines toward unity as the dark matter catches up to the galaxies (see Fig. 4).

2. The bias is scale dependent in the nonlinear regime, and this scale dependence itself depends on redshift. At  $z = 0$ , the galaxy correlation function is depressed below the dark matter correlation function at  $r \sim 0.5 h^{-1} \text{Mpc}$  and boosted above it at  $r \lesssim 0.1 h^{-1} \text{Mpc}$ . As a consequence, the galaxy correlation function is well described by a power law (of slope  $\gamma = 1.78 \pm 0.05$ ), while the dark matter correlation function is not (see Fig. 6). The dark matter correlation function is shallower at high redshift, while the galaxy correlation function slope is roughly constant, so the rms bias factors at 2 and  $8 h^{-1} \text{Mpc}$  start to diverge at  $z \gtrsim 1$  (see Fig. 9).

3. The correlation amplitude depends on galaxy baryon mass, with only a slight difference between the more and less massive halves of the complete sample (Fig. 7b), but significant increases of  $r_0$  and  $b_\sigma$  result if we select the 500 or 200 most massive galaxies in the simulation volume (Figs. 5 and 9; Table 2).

4. The dependence of the correlation amplitude on stellar population age is more striking than the dependence on baryon mass. The correlation functions of the older and younger halves of the complete sample have almost identical slopes, but their correlation lengths differ by nearly a factor of 2 (Fig. 7; Table 2).

5. At  $z = 0$ , the pairwise velocity dispersion  $\sigma_{12}(r)$  of galaxies is lower than that of the dark matter, by  $\sim 20\%$  (Fig. 8). There is a large difference between the pairwise dispersions of old and young galaxies, with the preferential location of old galaxies in dense environments making  $\sigma_{12}(r)$  about twice as high at  $r \lesssim 2 h^{-1} \text{Mpc}$ . The mean pairwise velocity  $v_{12}(r)$  is similar for galaxies and dark matter at  $z = 0$ , but the denser environments of older galaxies again cause a higher  $v_{12}(r)$ . At  $z = 1$ , the positive spatial bias of the galaxy population makes the mean pairwise velocity higher than that of the dark matter, although the pairwise dispersion remains somewhat lower.

6. The skewness of galaxy count fluctuations, as quantified by the hierarchical ratio  $S_{3f}$ , is lower than that of the dark matter by  $\sim 20\%$  (Fig. 10). The  $S_{3f}$  ratio and its scale dependence vary with galaxy baryon mass and age, in a fairly complex way.

7. The galaxy-mass cross-correlation function depends on galaxy mass and age in much the same way that the galaxy-galaxy correlation function itself does, although the trends are weaker because the mass distribution is the same in all cross-correlations (Fig. 11). For each class of galaxies, the bias function defined by the ratio  $\xi_{gg}(r)/\xi_{gm}(r)$  is roughly scale independent, even when it is significantly different from unity, although it can be somewhat different in amplitude and shape from the analogous bias function  $[\xi_{gg}(r)/\xi_{mm}(r)]^{1/2}$  (Fig. 12).

8. The average extended mass distributions around galaxies depend significantly on baryon mass and age (Fig. 13). The distribution around a typical high-mass galaxy falls more

steeply than an isothermal distribution, so that circular velocity curves  $[G\Delta M(r)/r]^{1/2}$  decline from  $r = 25 h^{-1}$  kpc out to  $r = 1 h^{-1}$  Mpc. For low-mass galaxies, on the other hand, the average mass profile is shallower than isothermal at  $r \gtrsim 100 h^{-1}$  kpc, where the individual galaxy halos begin to encounter the environment of a typical surrounding group. This effect is much more pronounced for old galaxies than for young galaxies. The ratio of mean aperture mass  $\Delta M(r)$  to galaxy baryon mass  $M_b$  is roughly independent of baryon mass for apertures ranging from  $r = 50 h^{-1}$  kpc to  $r = 1 h^{-1}$  Mpc, although the ratio is higher for the galaxies with the lowest  $M_b$ . The ratio is higher for old galaxies than for young galaxies, again demonstrating the preferential location of old galaxies in dense environments.

For the most part, the dark matter clustering of the  $\Lambda$ CDM model and these biases of the galaxy population lead to good qualitative and, where we have adequate statistics for comparison, quantitative agreement with observed galaxy clustering. The predicted galaxy correlation function is a power law of the observed slope and approximately the observed amplitude (see further discussion below). The redshift independence of the comoving correlation length agrees with measurements from Lyman break galaxy surveys at  $z \approx 3$  (Adelberger et al. 1998, 2003) and from deep redshift surveys extending to  $z \sim 1$  (Postman et al. 1998; Carlberg et al. 1999). The predicted dependence of clustering strength on baryon mass and population age agrees with the trends as a function of luminosity, color, and spectral type found in the 2dFGRS and SDSS (Norberg et al. 2001, 2002b; Zehavi et al. 2002). The predicted value of the hierarchical ratio  $S_{3f}$  agrees with measurements from angular clustering catalogs (Gaztañaga & Frieman 1994; Gaztañaga 2002; Szapudi et al. 2002). The value and scale independence of the bias inferred from the ratio of galaxy-galaxy and galaxy-mass correlation functions agrees with recent observational estimates (Hoekstra et al. 2001). The dependence of the galaxy-mass correlation function on population age parallels the dependence on galaxy type found by McKay et al. (2001), and the aperture mass  $\Delta M(260 h^{-1}$  kpc) is approximately proportional to galaxy baryon mass, as McKay et al. (2001) also find.

There are three quantitative discrepancies between our predictions and existing measurements, two fairly subtle and one more substantial. First, the power-law slope of  $\xi(r)$  for the  $L_*$  galaxy sample (the 500 most massive galaxies in the box at  $z = 0$ ) is slightly too steep,  $\gamma = 2.00 \pm 0.08$ , compared with observed values of 1.7–1.8, and there are hints of a departure from a power law at  $r \lesssim 1 h^{-1}$  Mpc. Second, the correlation length of this sample,  $r_0 = 4.5 \pm 0.4 h^{-1}$  Mpc, is low compared with the value  $r_0 = 6.3 \pm 0.8 h^{-1}$  Mpc measured for  $r$ -selected galaxies with  $L \approx L_*$  in the SDSS (Zehavi et al. 2002), although it is consistent with the value  $r_0 = 4.9 \pm 0.3 h^{-1}$  Mpc found for  $L_* b_J$ -selected galaxies in the 2dFGRS (Norberg et al. 2001). Because of the still limited size of our simulation volume, these discrepancies are of marginal statistical significance, and the  $N$ -body tests discussed in § 3.1 imply that the low correlation length, at least, is a consequence of the lower than average fluctuation amplitude in the particular realization of  $\Lambda$ CDM initial conditions used here.

The more serious quantitative discrepancy is the ratio of aperture mass  $\Delta M(260 h^{-1}$  kpc) to galaxy baryon mass  $M_b$ . For  $M_b \sim (0.5 - 1) \times 10^{11} M_\odot$ , the simulation prediction agrees with the result derived by combining the values of  $\Delta M/L$  of McKay et al. (2001) with the stellar mass-to-light ratios of Kauffmann et al. (2003), but at higher  $M_b$  the

predicted ratios are lower than the observationally inferred values by factors of  $\sim 1.5$ – $3$ , implying that the galaxy baryon masses in the simulation are too high (Fig. 14). As discussed in § 4, we believe that this discrepancy is largely numerical in origin, since our internal resolution tests (M. A. Fardal et al. 2004, in preparation) and comparisons between different SPH implementations (Springel & Hernquist 2002) suggest that in this mass range galaxy baryon masses are indeed overestimated relative to a simulation with higher resolution or more accurate treatment of the interface between hot and cold gas phases. A comparison between our best guess at resolution-corrected results and the McKay et al. (2001) data points yields closer agreement in the average values of  $\Delta M(260 h^{-1}$  kpc)/ $M_b$ , although the predicted ratios are still somewhat low. The comparison leaves limited room—at most a factor  $\sim 2$ —for additional astrophysical processes such as more aggressive feedback to substantially suppress gas cooling and star formation in this galaxy mass range. Consistent with this conclusion, Yang et al. (2003) find that the  $N$ -body+semianalytic models of Kauffman et al. (1999a), which do include much stronger stellar feedback, predict  $\Delta M(260 h^{-1}$  kpc)/ $L$  ratios that are too low (by a factor  $\sim 2$ ) relative to those of McKay et al. (2001). As emphasized by Guzik & Seljak (2002) and Seljak (2002a, 2002c), the constraints from galaxy-galaxy lensing measurements complement those from the galaxy luminosity function and the Tully-Fisher and fundamental plane relations because they suffer different systematic uncertainties and have different sensitivity to astrophysical and cosmological parameters. For example, lowering  $\Omega_m$  while keeping other parameters (including the linear power spectrum) fixed would tend to lower  $\Delta M(r)/L$  while having relatively little impact on the luminosity function (although galaxy formation, and hence galaxy luminosities, would still be affected by the different timescales and baryon-to-dark matter ratios of the lower  $\Omega_m$  model). The weak-lensing constraints become considerably more powerful as the precision and detail of the measurements improve.

When we have been able to compare, our results generally agree well with those of other hydrodynamic simulations (Pearce et al. 1999, 2001; Cen & Ostriker 2000; Yoshikawa et al. 2001), high-resolution  $N$ -body simulations that identify galaxies with subhalos (Colin et al. 1999; Kravtsov & Klypin 1999), and combined  $N$ -body+semianalytic models (Kauffmann et al. 1997, 1999a, 1999b; Governato et al. 1998; Benson et al. 2000a, 2000b, 2001). All these methods make qualitatively similar predictions for the evolution of the galaxy correlation function, the bias between galaxies and mass at the present day, the dependence of that bias on galaxy type, and the relative amplitude of dark matter and galaxy pairwise dispersions. This agreement suggests that these aspects of galaxy clustering and bias arise from robust aspects of the physics of galaxy formation that all these methods treat correctly, or at least similarly. Our conclusions about  $S_{3f}$  and galaxy-mass correlations are for the most part new, but we do not expect them to be fundamentally different from the predictions of other methods.

Returning to the main goals of the paper, we find that the combination of a  $\Lambda$ CDM cosmology with standard ideas about galaxy formation is, on the whole, remarkably successful at reproducing observed galaxy clustering and galaxy-galaxy lensing measurements. In terms of physical interpretation, the simulation analysis offers a number of insights. It shows that the detailed relation between the galaxy and dark halo populations can account for the difference between the predicted

correlation function of dark matter and the observed power-law form of the galaxy correlation function, and between the predicted pairwise velocity dispersion of the dark matter and the observed, lower dispersion of galaxies. It shows that the observed dependence of galaxy clustering on galaxy properties emerges naturally from the dependence of galaxy mass and population age on environment—basically, older and more massive galaxies formed in regions that collapsed early and are today biased with respect to the overall mass distribution. It shows that values of  $b_\sigma \sim 1$  and  $b_{\xi \times} \sim 1$  can emerge even when galaxies do not trace mass in detail. The quantitative discrepancies discussed above may also prove instructive, as the observations improve and the uncertainties in the simulations themselves are more thoroughly understood.

Our analysis makes a number of predictions that can be tested with future data or with future analyses of existing data. Our quantitative comparison with the observed type dependence of galaxy clustering has been limited by our finite simulation volume, which prevents us from predicting the clustering of rare classes of galaxies. However, it would be straightforward to test the quantitative predictions presented in, for example, Figure 7 by taking an observed sample above a luminosity threshold that yields the same space density as our complete sample and dividing it in two on the basis of luminosity or color (or, better still, on the basis of stellar mass and population age, using techniques such as those of Kauffmann et al. 2003). The predicted evolution of the mean pairwise velocity and pairwise dispersion (Fig. 8) can be tested with upcoming large redshift surveys such as the DEEP

and VIRMOS VLT programs. The systematic variations of  $S_{3\ell}$  with galaxy luminosity and age (Fig. 10) can be tested with the 2dFGRS and SDSS. The most informative comparisons will probably come from improved measurements of galaxy-galaxy lensing and cosmic shear, which will allow detailed tests of the trends predicted in Figures 12 and 13.

On all these fronts, observational efforts are advancing at a staggering pace. Comparisons between more comprehensive data and improved theoretical predictions over the next few years should tell us whether our understanding of dark matter clustering and the physics of galaxy formation is indeed complete, at least in terms of features that have major quantitative impact, or whether there are still important ingredients missing from our theoretical recipe.

We thank Andreas Berlind, Stephane Colombi, Mark Fardal, Henk Hoekstra, Guinevere Kauffmann, Tim McKay, Yannick Mellier, and Ludo van Waerbeke for valuable discussions on numerous aspects of this study. This work was supported by the NSF and by the NASA Astrophysical Theory Program and Long-Term Space Astrophysics Program. R. D. acknowledges the support of Hubble Fellowship grant HST-HF-01128.01-A from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555. D. H. W. acknowledges the hospitality of the Institute for Advanced Study and the Institut d’Astrophysique de Paris and the support of the Ambrose Monell Foundation and the French CNRS during the completion of this work.

#### REFERENCES

- Adelberger, K. L., Steidel, C. C., Giavalisco, M., Dickinson, M., Pettini, M., & Kellogg, M. 1998, *ApJ*, 505, 18
- Adelberger, K. L., Steidel, C. C., Shapley, A., & Pettini, M. 2003, *ApJ*, 584, 45
- Bahcall, N., Ostriker, J. P., Perlmutter, S., & Steinhardt, P. J. 1999, *Science*, 284, 1481
- Bahcall, N. A., et al. 2003, *ApJ*, 585, 182
- Baugh, C. M., Benson, A. J., Cole, S., Frenk, C. S., & Lacey, C. G. 1999, *MNRAS*, 305, L21
- Bean, A. J., Ellis, R. S., Shanks, T., Efstathiou, G., & Peterson, B. A. 1983, *MNRAS*, 205, 605
- Bennett, C. L., et al. 1996, *ApJ*, 464, L1
- Benson, A. J., Baugh, C. M., Cole, S., Frenk, C. S., & Lacey, C. G. 2000a, *MNRAS*, 316, 107
- Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M., & Lacey, C. G. 2000b, *MNRAS*, 311, 793
- Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., & Lacey, C. G. 2001, *MNRAS*, 327, 1041
- Berlind, A. A., & Weinberg, D. H. 2002, *ApJ*, 575, 587
- Berlind, A. A., et al. 2003, *ApJ*, 593, 1
- Blanton, M. R., et al. 2001, *AJ*, 121, 2358
- Bouchet, F. R., Strauss, M. A., Davis, M., Fisher, K. B., Yahil, A., & Huchra, J. P. 1993, *ApJ*, 417, 36
- Brainerd, T. G., Blandford, R. D., & Smail, I. 1996, *ApJ*, 466, 623
- Burles, S., & Tytler, D. 1997, *AJ*, 114, 1330
- . 1998, *ApJ*, 499, 699
- Carlberg, R., et al. 1999, *Philos. Trans. R. Soc. London, A*, 357, 167
- Carlberg, R. G., & Couchman, H. M. P. 1989, *ApJ*, 340, 47
- Cen, R., & Ostriker, J. P. 1992, *ApJ*, 399, L113
- . 2000, *ApJ*, 538, 83
- Chen, X., Weinberg, D. H., Katz, N., & Davé, R. 2003, *ApJ*, 594, 42
- Colin, P., Klypin, A. A., Kravtsov, A. V., & Khokhlov, A. M. 1999, *ApJ*, 523, 32
- Colombi, S., Szapudi, I., Jenkins, A., & Colberg, J. 2000, *MNRAS*, 313, 711
- Croft, R. A. C., Di Matteo, T., Davé, R., Hernquist, L., Katz, N., Fardal, M., & Weinberg, D. H. 2001, *ApJ*, 557, 67
- Croft, R. A. C., Weinberg, D. H., Bolte, M., Burles, S., Hernquist, L., Katz, N., Kirkman, D., & Tytler, D. 2002, *ApJ*, 581, 20
- Croft, R. A. C., Weinberg, D. H., Pettini, M., Katz, N., & Hernquist, L. 1999, *ApJ*, 520, 1
- Davé, R., Dubinski, J., & Hernquist, L. 1997, *NewA*, 2, 277
- Davé, R., Katz, N., & Weinberg, D. H. 2002, *ApJ*, 579, 23
- Davé, R., et al. 2001, *ApJ*, 552, 473
- Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, *ApJ*, 292, 371
- Davis, M., & Geller, M. J. 1976, *ApJ*, 208, 13
- Davis, M., Groth, E. J., & Peebles, P. J. E. 1977, *ApJ*, 212, L107
- Davis, M., & Peebles, P. J. E. 1983, *ApJ*, 267, 465
- dell’Antonio, I. P., & Tyson, J. A. 1996, *ApJ*, 473, L17
- Dressler, A. 1980, *ApJ*, 236, 351
- Efstathiou, G., Kaiser, N., Saunders, W., Lawrence, A., Rowan-Robinson, M., Ellis, R. S., & Frenk, C. S. 1990, *MNRAS*, 247, 10P
- Eke, V., Cole, S., & Frenk, C. S. 1996, *MNRAS*, 282, 263
- Evrard, A. E., Summers, F. J., & Davis, M. 1994, *ApJ*, 422, 11
- Fardal, M. A., Katz, N., Gardner, J. P., Hernquist, L., Weinberg, D. H., & Davé, R. 2001, *ApJ*, 562, 605
- Fardal, M. A., Katz, N., Weinberg, D. H., Davé, R., & Hernquist, L. 2002, *ApJ*, submitted (astro-ph/0107290)
- Ferreira, P. G., Juszkiewicz, R., Feldman, H. A., Davis, M., & Jaffe, A. H. 1999, *ApJ*, 515, L1
- Fischer, P., et al. 2000, *AJ*, 120, 1198
- Fisher, K. B. 1995, *ApJ*, 448, 494
- Fisher, K. B., Davis, M., Strauss, M. A., Yahil, A., & Huchra, J. P. 1994, *MNRAS*, 267, 927
- Fry, J. N. 1996, *ApJ*, 461, L65
- Fry, J. N., & Gaztañaga, E. 1993, *ApJ*, 413, 447
- Fry, J. N., & Melott, A. L. 1985, *ApJ*, 292, 395
- Gaztañaga, E. 2002, *ApJ*, 580, 144
- Gaztañaga, E., & Frieman, J. A. 1994, *ApJ*, 437, L13
- Gott, J. R., & Rees, M. J. 1975, *A&A*, 45, 365
- Gott, J. R., & Turner, E. L. 1979, *ApJ*, 232, L79
- Gott, J. R., Turner, E. L., & Aarseth, S. J. 1979, *ApJ*, 234, 13
- Gott, J. R., et al. 1989, *ApJ*, 340, 625
- Gott, J. R. I., & Turner, E. L. 1977, *ApJ*, 216, 357
- Governato, F., Baugh, C. M., Frenk, C. S., Cole, S., Lacey, C. G., Quinn, T. R., & Stadel, J. 1998, *Nature*, 392, 359

- Griffiths, R. E., Casertano, S., Im, M., & Ratnatunga, K. U. 1996, *MNRAS*, 282, 1159
- Guzik, J., & Seljak, U. 2001, *MNRAS*, 321, 439
- . 2002, *MNRAS*, 335, 311
- Guzzo, L., Strauss, M. A., Fisher, K. B., Giovanelli, R., & Haynes, M. P. 1997, *ApJ*, 489, 37
- Guzzo, L., et al. 2000, *A&A*, 355, 1
- Hamilton, A. J. S. 1988, *ApJ*, 331, L59
- . 1992, *ApJ*, 385, L5
- Hamilton, A. J. S., Matthews, A., Kumar, P., & Lu, E. 1991, *ApJ*, 374, L1
- Hatton, S., Devriendt, J. E. G., Ninin, S., Bouchet, F. R., Guiderdoni, B., & Vibert, D. 2003, *MNRAS*, 343, 75
- Hawkins, E., et al. 2004, *MNRAS*, in press (astro-ph/0212375)
- Hernquist, L., & Katz, N. 1989, *ApJS*, 70, 419
- Hoekstra, H., Yee, H. K. C., & Gladders, M. D. 2001, *ApJ*, 558, L11
- . 2002, *ApJ*, 577, 595
- Hudson, M. J., Gwyn, S. D. J., Dahle, H., & Kaiser, N. 1998, *ApJ*, 503, 531
- Jaffe, A., et al. 2001, *Phys. Rev. Lett.*, 86, 3475
- Jenkins, A., et al. 1998, *ApJ*, 499, 20
- Jing, Y. P., Börner, G., & Suto, Y. 2002, *ApJ*, 564, 15
- Jing, Y. P., Mo, H. J., & Börner, G. 1998, *ApJ*, 494, 1
- Juszkiewicz, R., Bouchet, F. R., & Colombi, S. 1993, *ApJ*, 412, L9
- Juszkiewicz, R., Weinberg, D. H., Amsterdamski, P., Chodorowski, M., & Bouchet, F. R. 1995, *ApJ*, 442, 39
- Kaiser, N. 1984, *ApJ*, 284, L9
- . 1987, *MNRAS*, 227, 1
- Katz, N., Hernquist, L., & Weinberg, D. H. 1992, *ApJ*, 399, L109
- . 1999, *ApJ*, 523, 463
- Katz, N., Weinberg, D. H., & Hernquist, L. 1996, *ApJS*, 105, 19 (KWH)
- Kauffmann, G., Colberg, J. M., Diaferio, A., & White, S. D. M. 1999a, *MNRAS*, 303, 188
- . 1999b, *MNRAS*, 307, 529
- Kauffmann, G., Nusser, A., & Steinmetz, M. 1997, *MNRAS*, 286, 795
- Kauffmann, G., et al. 2003, *MNRAS*, 341, 33
- Kollmeier, J. A., Weinberg, D. H., Davé, R., & Katz, N. 2003, *ApJ*, 594, 75
- Kravtsov, A. V., Berlind, A. A., Wechsler, R. H., Klypin, A. A., Gottloeber, S., Allgood, B., & Primack, J. R. 2003, *ApJ*, submitted (astro-ph/0308519)
- Kravtsov, A. V., & Klypin, A. 1999, *ApJ*, 520, 437
- Kroupa, P. 2002, *Science*, 295, 82
- Loveday, J., Maddox, S. J., Efstathiou, G., & Peterson, B. A. 1995, *ApJ*, 442, 457
- Ma, C., & Fry, J. N. 2000, *ApJ*, 543, 503
- Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, *MNRAS*, 242, P43
- Mann, R. G., Peacock, J. A., & Heavens, A. F. 1998, *MNRAS*, 293, 209
- Marzke, R. O., Geller, M. J., da Costa, L. N., & Huchra, J. P. 1995, *AJ*, 110, 477
- McDonald, P., Miralda-Escudé, J., Rauch, M., Sargent, W. L. W., Barlow, T. A., Cen, R., & Ostriker, J. P. 2000, *ApJ*, 543, 1
- McKay, T. A., et al. 2001, preprint (astro-ph/0108013)
- Mellier, Y., van Waerbeke, L., Maoli, R., Schneider, P., Jain, B., Bernardeau, F., Erben, T., & Fort, B. 2001, in *Proc. ESO/ECF/STScI Workshop, Deep Fields*, ed. S. Cristiani, A. Renzini, & R. E. Williams (Berlin: Springer), 252
- Mo, H. J., Jing, Y. P., & Börner, G. 1993, *MNRAS*, 264, 825
- Murali, C., Katz, N., Hernquist, L., Weinberg, D. H., & Davé, R. 2002, *ApJ*, 571, 1
- Narayanan, V. K., Berlind, A. A., & Weinberg, D. H. 2000, *ApJ*, 528, 1
- Norberg, P., et al. 2001, *MNRAS*, 328, 64
- Norberg, P., et al. 2002a, *MNRAS*, 336, 907
- . 2002b, *MNRAS*, 332, 827
- Park, C. 1990, *MNRAS*, 242, P59
- Park, C., Vogeley, M. S., Geller, M. J., & Huchra, J. P. 1994, *ApJ*, 431, 569
- Peacock, J. A., & Dodds, S. J. 1996, *MNRAS*, 280, L19
- Peacock, J. A., & Smith, R. E. 2000, *MNRAS*, 318, 1144
- Pearce, F. R., Jenkins, A., Frenk, C. S., White, S. D. M., Thomas, P. A., Couchman, H. M. P., Peacock, J. A., & Efstathiou, G. 2001, *MNRAS*, 326, 649
- Pearce, F. R., et al. 1999, *ApJ*, 521, L99
- Peebles, P. J. E. 1974, *A&A*, 32, 197
- . 1980, *The Large Scale Structure of the Universe* (Princeton: Princeton Univ. Press)
- Percival, W. J., et al. 2002, *MNRAS*, 337, 1068
- Perlmutter, S., et al. 1999, *ApJ*, 517, 565
- Postman, M., & Geller, M. J. 1984, *ApJ*, 281, 95
- Postman, M., Lauer, T. R., Szapudi, I., & Oegerle, W. 1998, *ApJ*, 506, 33
- Quinn, T., Katz, N., Stadel, J., & Lake, G. 1997, preprint (astro-ph/9710043)
- Ratcliffe, A., Shanks, T., Parker, Q. A., & Fong, R. 1998, *MNRAS*, 296, 173
- Riess, A., et al. 1998, *AJ*, 116, 1009
- Schechter, P. 1976, *ApJ*, 203, 297
- Scoccimarro, R., Sheth, R. K., Hui, L., & Jain, B. 2001, *ApJ*, 546, 20
- Seljak, U. 2000, *MNRAS*, 318, 203
- . 2002a, *MNRAS*, 334, 797
- . 2002b, *MNRAS*, 337, 769
- . 2002c, *MNRAS*, 337, 774
- Sheth, R. K., Hui, L., Diaferio, A., & Scoccimarro, R. 2001, *MNRAS*, 326, 463
- Smith, D. R., Bernstein, G. M., Fischer, P., & Jarvis, M. 2001, *ApJ*, 551, 643
- Somerville, R. S., Lemson, G., Sigad, Y., Dekel, A., Kauffmann, G., & White, S. D. M. 2001, *MNRAS*, 320, 289
- Spergel, D. N., et al. 2003, *ApJS*, 148, 175
- Springel, V., & Hernquist, L. 2002, *MNRAS*, 333, 649
- Strauss, M. A., et al. 2002, *AJ*, 124, 1810
- Szapudi, I., & Szalay, A. S. 1993, *ApJ*, 408, 43
- Szapudi, I., et al. 2002, *ApJ*, 570, 75
- Tegmark, M., & Bromley, B. C. 1999, *ApJ*, 518, L69
- Tegmark, M., & Peebles, P. J. E. 1998, *ApJ*, 500, L79
- Totsuji, H., & Kihara, T. 1969, *PASJ*, 21, 221
- Tully, R. B., & Fisher, J. R. 1977, *A&A*, 54, 661
- Tyson, J. A., Valdes, F., Jarvis, J. F., & Mills, A. P. 1984, *ApJ*, 281, L59
- Tytler, D., Fan, X. M., & Burles, S. 1996, *Nature*, 381, 207
- Weinberg, D. H., & Cole, S. 1992, *MNRAS*, 259, 652
- Weinberg, D. H., Davé, R., Gardner, J. P., Hernquist, L., & Katz, N. 1999, in *ASP Conf. Ser. 191, Photometric Redshifts and High Redshift Galaxies*, ed. R. Weymann, L. Storrie-Lombardi, M. Sawicki, & R. Brunner (ASP: San Francisco), 341
- Weinberg, D. H., Hernquist, L., & Katz, N. 1997, *ApJ*, 477, 8
- . 2002, *ApJ*, 571, 15
- White, M., Hernquist, L., & Springel, V. 2001, preprint (astro-ph/0107023)
- White, S. D. M., Frenk, C. S., & Davis, M. 1983, *ApJ*, 274, L1
- Willmer, C. N. A., da Costa, L. N., & Pellegrini, P. S. 1998, *AJ*, 115, 869
- Wilson, G., Kaiser, N., & Luppino, G. A. 2001a, *ApJ*, 556, 601
- Wilson, G., Kaiser, N., Luppino, G. A., & Cowie, L. L. 2001b, *ApJ*, 555, 572
- Yang, X., Mo, H. J., Kauffmann, G., & Chu, Y. 2003, *MNRAS*, 339, 387
- Yoshikawa, K., Taruya, A., Jing, Y. P., & Suto, Y. 2001, *ApJ*, 558, 520
- Zehavi, I., et al. 2002, *ApJ*, 571, 172
- . 2003, *ApJ*, submitted