ARTICLES

Simulations of the formation, evolution and clustering of galaxies and quasars

Volker Springel[1], Simon D. M. White[1], Adrian Jenkins[2], Carlos S. Frenk[2], Naoki Yoshida[3], Liang Gao[1], Julio Navarro[4], Robert Thacker[5], Darren Croton[1], John Helly[2], John A. Peacock[6], Shaun Cole[2], Peter Thomas[7], Hugh Couchman[5], August Evrard[8], Jörg Colberg[9] & Frazer Pearce[10]

The cold dark matter model has become the leading theoretical picture for the formation of structure in the Universe. This model, together with the theory of cosmic inflation, makes a clear prediction for the initial conditions for structure formation and predicts that structures grow hierarchically through gravitational instability. Testing this model requires that the precise measurements delivered by galaxy surveys can be compared to robust and equally precise theoretical calculations. Here we present a simulation of the growth of dark matter structure using $2{,}160^3$ particles, following them from redshift $z = 127$ to the present in a cube-shaped region 2,230 billion lightyears on a side. In postprocessing, we also follow the formation and evolution of the galaxies and quasars. We show that baryon-induced features in the initial conditions of the Universe are reflected in distorted form in the low-redshift galaxy distribution, an effect that can be used to constrain the nature of dark energy with future generations of observational surveys of galaxies.

Recent large surveys, such as the 2-degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), have characterized, much more accurately than before, both the spatial clustering and the physical properties of low-redshift galaxies. Major ongoing campaigns exploit the new generation of 8-m-class telescopes and the Hubble Space Telescope to acquire data of comparable quality at high redshift. Other surveys target the weak image shear caused by gravitational lensing to extract precise measurements of the distribution of dark matter around galaxies and galaxy clusters. The principal goals of all these surveys are to shed light on how galaxies form, to test the current model for the growth of cosmic structure, and to search for signatures that may clarify the nature of dark matter and dark energy. These goals can be achieved only if the accurate measurements delivered by the surveys can be compared to robust and equally precise theoretical predictions.

Two problems have so far precluded such predictions: (1) accurate estimates of clustering require simulations of extreme dynamic range, encompassing volumes large enough to contain representative populations of rare objects (such as rich galaxy clusters or quasars), yet resolving the formation of individual low-luminosity galaxies; (2) critical aspects of galaxy-formation physics are uncertain and beyond the reach of direct simulation. Such aspects include, for example, the structure of the interstellar medium and its consequences for star formation and for the generation of galactic winds, the ejection and mixing of heavy elements, and active galactic nuclei (AGN) feeding and feedback effects; these must be treated by phenomenological models whose form and parameters are adjusted by trial and error as part of the overall data-modelling process. Here we have developed a framework that combines very large computer simulations of structure formation with post-hoc modelling of galaxy-formation physics to offer a practical solution to these two entwined problems.

During the past two decades, the cold dark matter (CDM) model, augmented with a dark energy field (which may take the form of a cosmological constant, Λ), has been developed into the standard theoretical model for galaxy formation. It assumes that structure grew from weak density fluctuations present in the otherwise homogeneous and rapidly expanding early Universe. These fluctuations are amplified by gravity, eventually turning into the rich structure that we see around us today. Confidence in the validity of this model has been boosted by recent observations. Measurements of the cosmic microwave background (CMB) by the Wilkinson Microwave Anisotropy Probe (WMAP) satellite[1] were combined with the 2dFGRS to confirm the central tenets of the model and to allow an accurate determination of the geometry and matter content of the Universe about 380,000 yr after the Big Bang[2]. The data suggest that the early density fluctuations were a gaussian random field, as predicted by inflationary theory, and that the current energy density is dominated by some form of dark energy. This analysis is supported by the apparent acceleration of the current cosmic expansion inferred from studies of distant supernovae[3,4], as well as by the low matter density derived from the baryon fraction of clusters[5].

While the initial, linear growth of density perturbations can be calculated analytically, the collapse of fluctuations and the subsequent hierarchical build-up of structure is a highly nonlinear process that is accessible only through direct numerical simulation[6]. The dominant mass component, the cold dark matter, is assumed to be made of elementary particles that currently interact only gravitationally, so the collisionless dark matter fluid can be represented by a set of discrete point particles. This representation as an N-body system is a coarse approximation whose fidelity improves as the number of particles in the simulation increases.

The high-resolution simulation described here—dubbed the

[1]Max-Planck-Institute for Astrophysics, Karl-Schwarzschild-Strasse 1, 85740 Garching, Germany. [2]Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK. [3]Department of Physics, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan. [4]Department of Physics & Astronomy, University of Victoria, Victoria, British Columbia V8P 5C2, Canada. [5]Department of Physics & Astronomy, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4M1, Canada. [6]Institute of Astronomy, University of Edinburgh, Blackford Hill, Edinburgh EH9 3HJ, UK. [7]Department of Physics & Astronomy, University of Sussex, Falmer, Brighton BN1 9QH, UK. [8]Department of Physics & Astronomy, University of Michigan, Ann Arbor, Michigan 48109-1120, USA. [9]Department of Physics & Astronomy, University of Pittsburgh, 3941 O'Hara Street, Pittsburgh, Pennsylvania 15260, USA. [10]Physics and Astronomy Department, University of Nottingham, Nottingham NG7 2RD, UK.

Millennium Simulation because of its size—was carried out by the Virgo Consortium, a collaboration of British, German, Canadian and US astrophysicists. It follows $N = 2{,}160^3 \cong 1.0078 \times 10^{10}$ particles from redshift $z = 127$ to the present in a cubic region $500h^{-1}$ Mpc on a side, where $1 + z$ is the expansion factor of the Universe relative to the present and h is Hubble's constant in units of 100 km s^{-1} Mpc^{-1}. With ten times as many particles as the previous largest computations of this kind[7–9] (see Supplementary Information), it offers substantially improved spatial and time resolution within a large cosmological volume. Combining this simulation with new techniques for following the formation and evolution of galaxies, we predict the positions, velocities and intrinsic properties of all galaxies brighter than the Small Magellanic Cloud throughout volumes comparable to the largest current surveys. Crucially, this also allows us to establish evolutionary links between objects observed at different epochs. For example, we demonstrate that galaxies with supermassive central black holes can plausibly form early enough in the standard cold dark matter cosmology to host the first known quasars, and that these end up at the centres of rich galaxy clusters today.

Dark matter haloes and galaxies

The mass distribution in a ΛCDM universe has a complex topology, often described as a 'cosmic web'[10]. This is visible in Fig. 1 (see also the corresponding Supplementary Video). The zoomed-out panel at the bottom of the figure reveals a tight network of cold dark matter clusters and filaments of characteristic size $\sim 100h^{-1}$ Mpc. On larger scales, there is little discernible structure and the distribution appears homogeneous and isotropic. Subsequent images zoom in by factors of four onto the region surrounding one of the many rich galaxy clusters. The final image reveals several hundred dark matter substructures, resolved as independent, gravitationally bound objects orbiting within the cluster halo. These substructures are the remnants of dark matter haloes that fell into the cluster at earlier times.

The space density of dark matter haloes at various epochs in the simulation is shown in Fig. 2. At present, there are about 18 million haloes above a detection threshold of 20 particles; 49.6% of all particles are included in these haloes. These statistics provide the most precise determination to date of the mass function of cold dark matter haloes[11,12]. In the range that is well sampled in our simulation ($z \leq 12$, $M \geq 1.7 \times 10^{10} h^{-1} M_{\odot}$, where $M_{\odot}$ is the solar mass), our results are remarkably well described by the analytic formula proposed by ref. 11 from fits to previous simulations. Theoretical models based on an ellipsoidal excursion set formulation[13] give a less accurate, but still reasonable, match. However, the commonly used Press–Schechter formula[14] underpredicts the high-mass end of the mass function by up to an order of magnitude. Previous studies of the abundance of rare objects, such as luminous quasars or clusters,

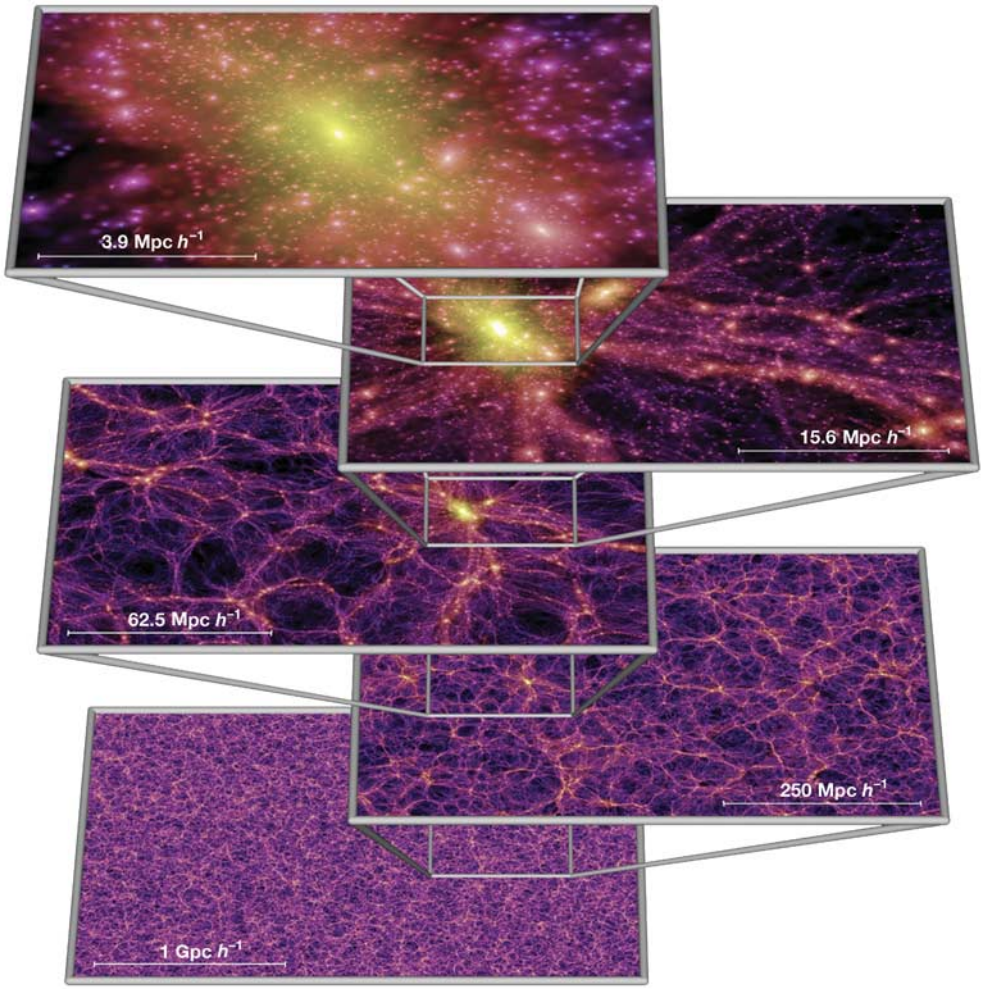

Figure 1 | The dark matter density field on various scales. Each individual image shows the projected dark matter density field in a slab of thickness $15h^{-1}$ Mpc (sliced from the periodic simulation volume at an angle chosen to avoid replicating structures in the lower two images), colour-coded by density and local dark matter velocity dispersion. The zoom sequence displays consecutive enlargements by factors of four, centred on one of the many galaxy cluster haloes present in the simulation.

based on this formula may contain large errors[15]. We return below to the important question of the abundance of quasars at early times.

To track the formation of galaxies and quasars in the simulation, we implement a semi-analytic model to follow gas, star and supermassive black-hole processes within the merger history trees of dark matter haloes and their substructures (see Supplementary Information). The trees contain a total of about 800 million nodes, each corresponding to a dark matter subhalo and its associated galaxies. This methodology allows us to test, during postprocessing, many different phenomenological treatments of gas cooling, star formation, AGN growth, feedback, chemical enrichment and so on. Here, we use an update of models described in refs 16 and 17, which are similar in spirit to previous semi-analytic models[18–23]; the modelling assumptions and parameters are adjusted by trial and error to fit the observed properties of low-redshift galaxies, primarily their joint luminosity–colour distribution and their distributions of morphology, gas content and central black-hole mass. Our use of a high-resolution simulation, particularly our ability to track the evolution of dark matter substructures, removes much of the uncertainty of the more traditional semi-analytic approaches based on Monte Carlo realizations of merger trees. Our technique provides accurate positions and peculiar velocities for all the model galaxies. It also enables us to follow the evolutionary history of individual objects and thus to investigate the relationship between populations seen at different epochs. It is the ability to establish such evolutionary connections that makes this kind of modelling so powerful for interpreting observational data.

The fate of the first quasars

Quasars are among the most luminous objects in the Universe and can be detected at huge cosmological distances. Their luminosity is thought to be powered by accretion onto a central, supermassive black hole. Bright quasars have now been discovered as far back as redshift $z = 6.43$ (ref. 24), and are believed to harbour central black holes with a mass a billion times that of the Sun. At redshift $z \approx 6$, their co-moving space density is estimated to be $\sim(2.2 \pm 0.73) \times 10^{-9} h^3\,\mathrm{Mpc}^{-3}$ (ref. 25). Whether such extremely rare objects can form at all in a ΛCDM cosmology is unknown.

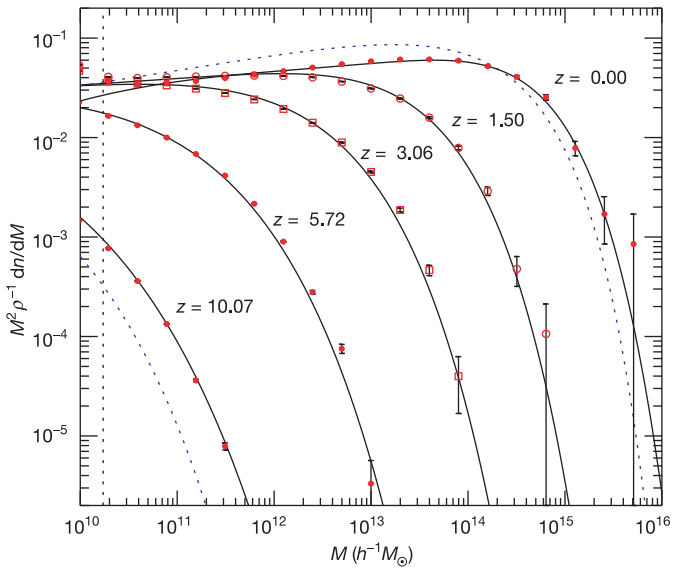

Figure 2 | Differential halo number density as a function of mass and epoch. The function $n(M, z)$ gives the co-moving number density of haloes less massive than M. We plot it as the halo multiplicity function $M^2\rho^{-1}\mathrm{d}n/\mathrm{d}M$ (symbols with 1-σ error bars), where ρ is the mean density of the Universe. Groups of particles were found using a friends-of-friends algorithm[6] with linking length equal to 0.2 of the mean particle separation. The fraction of mass bound to haloes of more than 20 particles (vertical dotted line) grows from 6.42×10^{-4} at $z = 10.07$ to 0.496 at $z = 0$. Solid lines are predictions from an analytic fitting function proposed in previous work[11], and the dashed blue lines give the Press–Schechter model[14] at $z = 10.07$ and $z = 0$.

A volume the size of the Millennium Simulation should contain, on average, just under one quasar at the above space density. Just what sort of object should be associated with these 'first quasars' is, however, a matter of debate. In the local Universe, it appears that every bright galaxy hosts a supermassive black hole and there is a remarkably good correlation between the mass of the central black hole and the stellar mass or velocity dispersion of the bulge of the host galaxy[26]. It would therefore seem natural to assume that, at any epoch, the brightest quasars are always hosted by the largest galaxies. In our simulation, 'large galaxies' can be identified in various ways, for example, according to their dark matter halo mass, stellar mass or instantaneous star-formation rate. We have identified the ten 'largest' objects defined in these three ways at redshift $z = 6.2$. It turns out that these criteria all select essentially the same objects: the eight largest galaxies by halo mass are identical to the eight largest galaxies by stellar mass; only the ranking differs. Somewhat larger differences are present when galaxies are selected by star-formation rate, but the four first-ranked galaxies are still among the eight identified according to the other two criteria.

In Fig. 3, we illustrate the environment of a 'first quasar' candidate in our simulation at $z = 6.2$. The object lies on one of the most prominent dark matter filaments and is surrounded by a large number of other, much fainter galaxies. It has a stellar mass of $6.8 \times 10^{10} h^{-1} M_\odot$, the largest in the entire simulation at $z = 6.2$, a dark matter virial mass of $3.9 \times 10^{12} h^{-1} M_\odot$, and a star-formation rate of $235 M_\odot\,\mathrm{yr}^{-1}$. In the local Universe, central black-hole masses are typically ~1/1,000 of the bulge stellar mass[27], but in the model we test here these massive early galaxies have black-hole masses in the range 10^8–$10^9 M_\odot$, significantly larger than low-redshift galaxies of similar stellar mass. To attain the observed luminosities, they must convert infalling mass to radiated energy with a somewhat higher efficiency than the $\sim 0.1c^2$ expected for accretion onto a non-spinning black hole (where c is the speed of light in vacuum).

Within our simulation we can readily address fundamental questions such as: Where are the descendants of the early quasars today? What were their progenitors? By tracking the merging history trees of the host haloes, we find that all our quasar candidates end up today as central galaxies in rich clusters. For example, the object depicted in Fig. 3 lies, today, at the centre of the ninth most massive cluster in the volume, of mass $M = 1.46 \times 10^{15} h^{-1} M_\odot$. The candidate with the largest virial mass at $z = 6.2$ (which has stellar mass $4.7 \times 10^{10} h^{-1} M_\odot$, virial mass $4.85 \times 10^{12} h^{-1} M_\odot$, and star-formation rate $218 M_\odot\,\mathrm{yr}^{-1}$) ends up in the second most massive cluster, of mass $3.39 \times 10^{15} h^{-1} M_\odot$. Following the merging tree backwards in time, we can trace our quasar candidate back to redshift $z = 16.7$, when its host halo had a mass of only $1.8 \times 10^{10} h^{-1} M_\odot$. At this epoch, it is one of just 18 objects that we identify as collapsed systems with ≥20 particles. These results confirm the view that rich galaxy clusters are rather special places. Not only are they the largest virialized structures today, they also lie in the regions where the first structures developed at high redshift. Thus, the best place to search for the oldest stars in the Universe or for the descendants of the first supermassive black holes is at the centres of present-day rich galaxy clusters.

The clustering evolution of dark matter and galaxies

The combination of a large-volume, high-resolution N-body simulation with realistic modelling of galaxies enables us to make precise theoretical predictions for the clustering of galaxies as a function of redshift and intrinsic galaxy properties. These can be compared directly with existing and planned surveys. The two-point correlation function of our model galaxies at redshift $z = 0$ is plotted in Fig. 4 and is compared with a recent measurement from the 2dFGRS

(ref. 28). The prediction is remarkably close to a power law, confirming with much higher precision the results of earlier semi-analytic[23,29] and hydrodynamic[30] simulations. This precision will allow interpretation of the small but measurable deviations from a pure power law found in the most recent data[31,32]. The simple power-law form contrasts with the more complex behaviour exhibited by the dark matter correlation function, but is really no more than a coincidence. Correlation functions for galaxy samples with different selection criteria or at different redshifts do not, in general, follow power laws.

Although our semi-analytic model was not tuned to match observations of galaxy clustering, it not only produces the excellent overall agreement shown in Fig. 4, but also reproduces the observed dependence of clustering on magnitude and colour in the 2dFGRS and SDSS (refs 33–35), as shown in Fig. 5. The agreement is particularly good for the dependence of clustering on luminosity. The colour dependence of the slope is matched precisely, but the amplitude difference is greater in our model than is observed[35]. Note that our predictions for galaxy correlations split by colour deviate substantially from power laws. Such predictions can be easily tested against survey data to clarify the physical processes responsible for the observed variations.

In contrast to the near-power-law behaviour of galaxy correlations on small scales, the large-scale clustering pattern may show interesting structure. Coherent oscillations in the primordial plasma give rise to the well-known acoustic peaks in the CMB (refs 2, 36, 37) and also leave an imprint in the linear power spectrum of the dark matter. Detection of these 'baryon wiggles' would not only provide an excellent consistency check for the cosmological model, but could also have important practical applications. The characteristic scale of the wiggles provides a 'standard ruler' that may be used to constrain the equation of state of the dark energy[38]. A critical question when designing future surveys is whether these baryon wiggles are present and are detectable in the galaxy distribution, particularly at high redshift.

On large scales and at early times, the mode amplitudes of the dark matter power spectrum grow linearly, roughly in proportion to the cosmological expansion factor. Nonlinear evolution accelerates the growth on small scales when the dimensionless power $\Delta^2(k) = k^3P(k)/(2\pi^2)$ approaches unity (the power spectrum $P(k)$ measures the variance of density fluctuations on the scale of wavenumber k); this regime can only be studied accurately using numerical simulations. In the Millennium Simulation, we are able to determine the nonlinear power spectrum over a larger range of scales

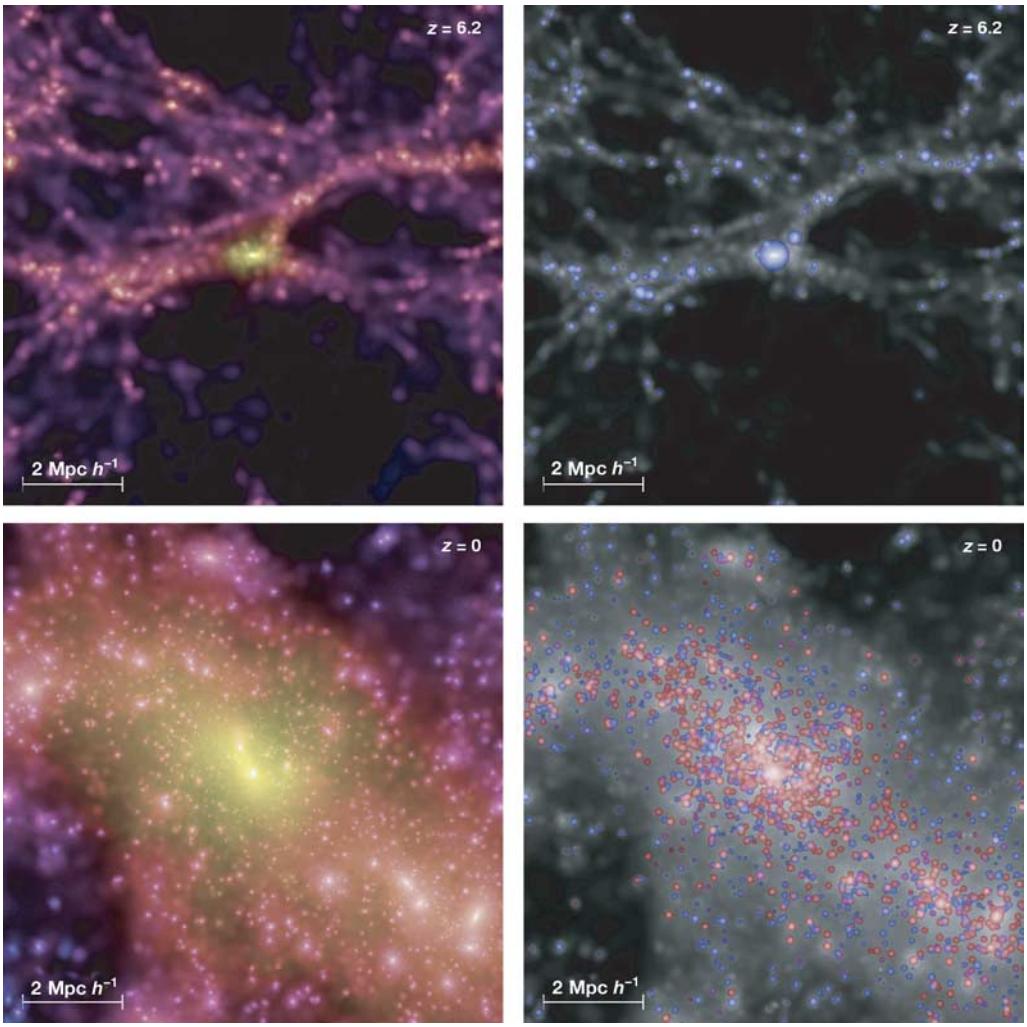

Figure 3 | Environment of a 'first quasar candidate' at high and low redshifts. The two panels on the left show the projected dark matter distribution in a cube of co-moving sidelength $10h^{-1}$ Mpc, colour-coded according to density and local dark matter velocity dispersion. The panels on the right show the galaxies of the semi-analytic model overlaid on a greyscale image of the dark matter density. The volume of the sphere representing each galaxy is proportional to its stellar mass, and the chosen colours encode the restframe stellar B–V colour index. While at $z = 6.2$ (top panels) all galaxies appear blue owing to ongoing star formation, many of the galaxies that have fallen into the rich cluster at $z = 0$ (bottom) have turned red.

than was possible in earlier work[39]: almost five orders of magnitude in wavenumber k.

At present, the acoustic oscillations in the matter power spectrum are expected to fall in the transition region between linear and nonlinear scales. In Fig. 6, we examine the matter power spectrum in our simulation in the region of the oscillations. Dividing by the smooth power spectrum of a ΛCDM model with no baryons[40] highlights the baryonic features in the initial power spectrum of the simulation, although there is substantial scatter owing to the small number of large-scale modes. Because linear growth preserves the relative mode amplitudes, we can approximately correct for this scatter by scaling the measured power in each bin by a multiplicative factor based on the initial difference between the actual bin power and the mean power expected in our ΛCDM model. This makes the effects of nonlinear evolution on the baryon oscillations more clearly visible. As Fig. 6 shows, nonlinear evolution not only accelerates growth but also reduces the baryon oscillations: scales near peaks grow slightly more slowly than scales near troughs. This is a consequence of the mode–mode coupling characteristic of nonlinear growth. In spite of these effects, the first two 'acoustic peaks' (at $k \approx 0.07$ and $k \approx 0.13h\,\mathrm{Mpc}^{-1}$, respectively) in the dark matter distribution do survive in distorted form until the present day (see Fig. 6f).

Are the baryon wiggles also present in the galaxy distribution? Fig. 6 shows that the answer to this important question is "yes". The $z = 0$ panel (Fig. 6f) shows the power spectrum for all model galaxies brighter than $M_{\mathrm{B}} = -17$ (M_{B} is the magnitude in the optical blue waveband). On the largest scales, the galaxy power spectrum has the same shape as that of the dark matter, but with slightly lower amplitude corresponding to an 'antibias' of 8%. Samples of brighter galaxies show less antibias, while for the brightest galaxies, the bias becomes slightly positive. Figure 6d, e also shows measurements of the power spectrum of luminous galaxies at redshifts $z = 0.98$ and $z = 3.06$. Galaxies at $z = 0.98$ were selected to have a magnitude $M_{\mathrm{B}} < -19$ in the restframe, whereas galaxies at $z = 3.06$ were selected to have stellar mass larger than $5.83 \times 10^{9} h^{-1} M_{\odot}$, corresponding to a space density of $8 \times 10^{-3} h^{3}\,\mathrm{Mpc}^{-3}$, similar to that of the Lyman-break galaxies observed at $z \approx 3$ (ref. 41). Signatures of the first two acoustic peaks are clearly visible at both redshifts, even though the density field of the $z = 3$ galaxies is much more strongly biased with respect to the dark matter (by a factor $b = 2.7$, where $b = [P_{\mathrm{gal}}(k)/P_{\mathrm{dm}}(k)]^{1/2}$) than at low redshift. Selecting galaxies by their star-formation rate rather than their stellar mass (above $10.6 M_{\odot}\,\mathrm{yr}^{-1}$ for an equal space density at $z = 3$) produces very similar results.

Our analysis demonstrates conclusively that baryon wiggles should indeed be present in the galaxy distribution out to redshift $z = 3$. This has been assumed, but not justified, in recent proposals to use evolution of the large-scale galaxy distribution to constrain the nature of the dark energy. To establish whether the baryon oscillations can be measured in practice with the requisite accuracy will require detailed modelling of the selection criteria of an actual survey, and a thorough understanding of the systematic effects that will inevitably be present in real data. These issues can only be properly addressed by means of specially designed mock

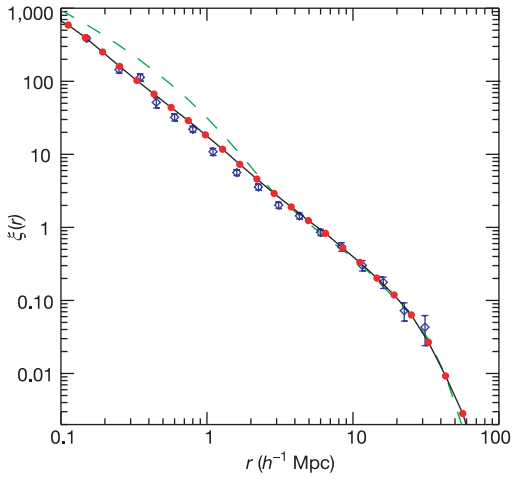

Figure 4 | Galaxy two-point correlation function, $\xi(r)$, at the present epoch as a function of separation r. Red symbols (with vanishingly small Poisson error bars) show measurements for model galaxies brighter than $M_K = -23$, where M_K is the magnitude in the K-band. Data for the large spectroscopic redshift survey 2dFGRS (ref. 28) are shown as blue diamonds together with their 1-σ error bars. The SDSS (ref. 34) and APM (ref. 31) surveys give similar results. Both for the observational data and for the simulated galaxies, the correlation function is very close to a power law for $r \leq 20h^{-1}$ Mpc. By contrast, the correlation function for the dark matter (dashed green line) deviates strongly from a power law.

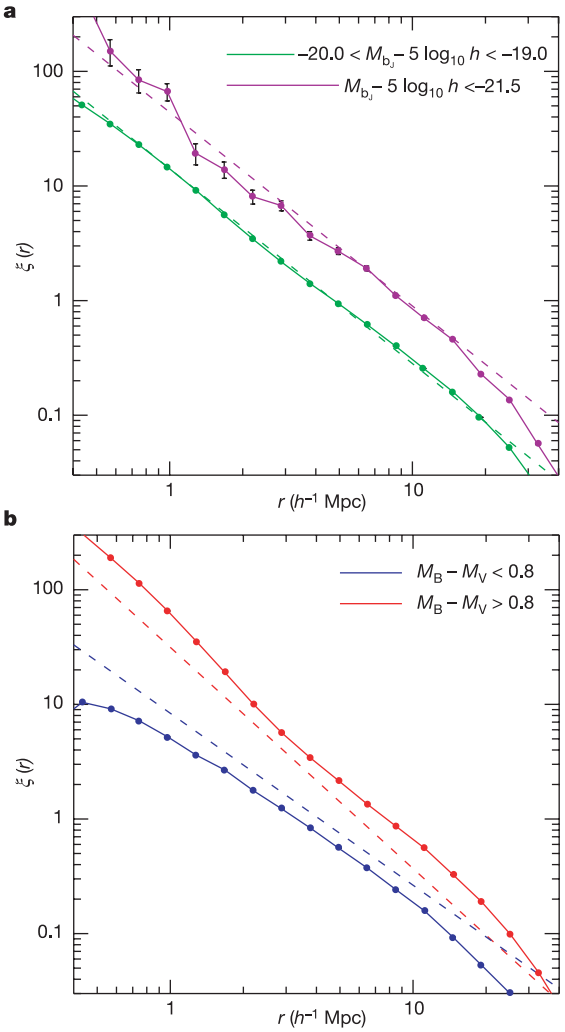

Figure 5 | Galaxy clustering as a function of luminosity and colour. a, the two-point correlation function of our galaxy catalogue at $z = 0$ split by luminosity in the b_J-band filter (symbols with 1-σ error bars) Brighter galaxies are more strongly clustered, in quantitative agreement with observations[33] (dashed lines). Splitting galaxies according to colour (**b**), we find that red galaxies are more strongly clustered with a steeper correlation slope than blue galaxies. Observations[35] (dashed lines) show a similar trend, although the difference in clustering amplitude is smaller than in this particular semi-analytic model.

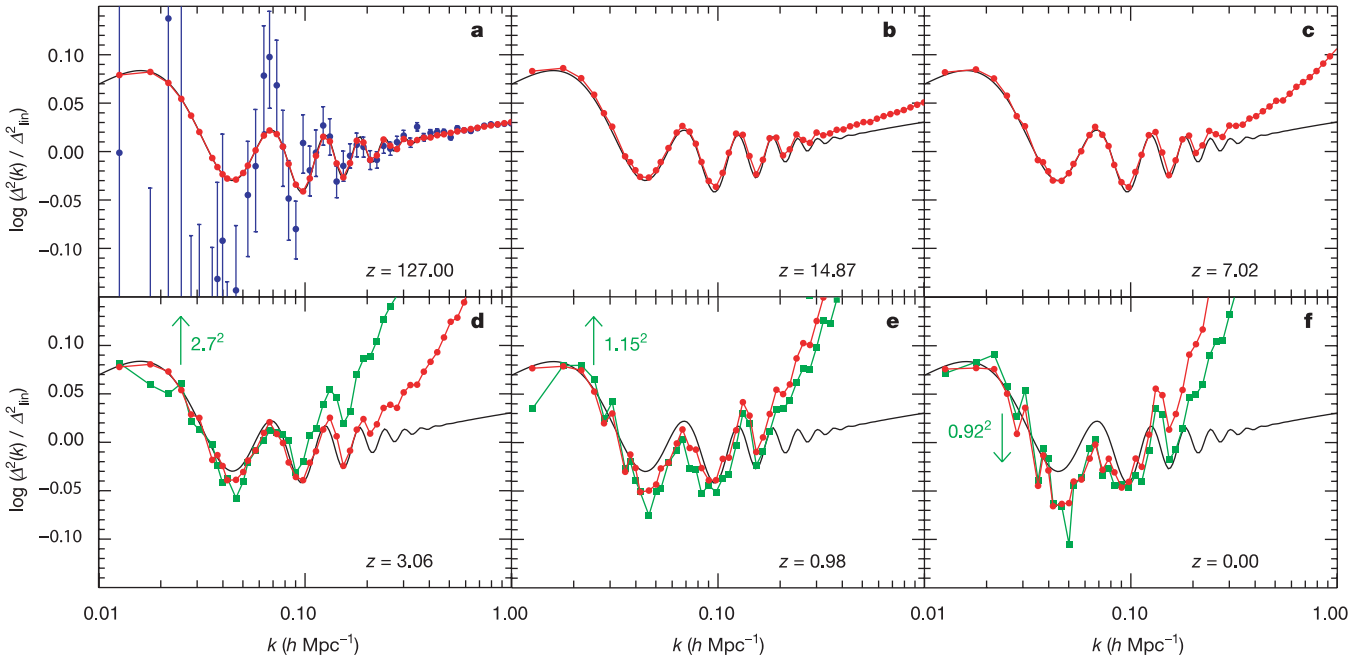

Figure 6 | Power spectra of the dark matter and galaxy distributions in the baryon oscillation region. a–f, All measurements have been divided by a linearly evolved, CDM-only power spectrum $\Delta^2_{\rm lin}$ (ref. 40). Red circles show the dark matter, and green squares the galaxies. Blue symbols give the actual realization of the initial fluctuations in our simulation, which scatters around the mean input power (black lines) owing to the finite number of modes (error bars give the 1-σ scatter around the mean power in each bin). Because linear growth preserves relative mode amplitudes, we correct the power in each bin to the expected input power and apply these scaling factors at all other times. **d**, At $z = 3.06$, galaxies with stellar mass above $5.83 \times 10^9 h^{-1} M_\odot$ and space-density of $8 \times 10^{-3} h^3\,{\rm Mpc}^{-3}$ were selected. Their large-scale density field is biased by a factor $b = 2.7$ with respect to the dark matter (the galaxy measurement has been divided by b^2). **f**, At $z = 0$, galaxies brighter than $M_{\rm B} = -17$ and a space density higher by a factor of $\sim$7.2 were selected. They exhibit a slight antibias of $b = 0.92$. **e**, The corresponding numbers for $z = 0.98$ are $M_{\rm B} = -19$ and $b = 1.15$.

catalogues constructed from realistic simulations. We plan to construct suitable mock catalogues from the Millennium Simulation and make them publicly available. Our provisional conclusion, however, is that the next generation of galaxy surveys offers excellent prospects for constraining the equation of state of the dark energy.

N-body simulations of CDM universes are now of such size and quality that realistic modelling of galaxy formation in volumes matched to modern surveys has become possible. Detailed studies of galaxy and AGN evolution exploiting the unique data set of the Millennium Simulation therefore make stringent new tests of the theory of hierarchical galaxy formation possible. Using the simulation, we demonstrate that quasars can plausibly form sufficiently early in a ΛCDM universe to be compatible with observation, that their progenitors were already massive by $z \approx 16$, and that their $z = 0$ descendants lie at the centres of cD galaxies in rich galaxy clusters. Interesting tests of our predictions will become possible if observations of the black-hole demographics can be extended to high redshift, allowing, for example, a measurement of the evolution of the relationship between supermassive black-hole masses and the velocity dispersion of their host stellar bulges.

We have also demonstrated that a power-law galaxy autocorrelation function can arise naturally in a ΛCDM universe, but that this behaviour has no simple physical cause and is merely a coincidence. Galaxy surveys will soon reach sufficient statistical power to measure precise deviations from power laws for galaxy subsamples, and we expect that comparisons of the kind we have illustrated will lead to tight constraints on the physical processes included in the galaxy-formation modelling. Finally, we have demonstrated that the baryon-induced oscillations recently detected in the CMB power spectrum should survive in distorted form not only in the nonlinear dark matter power spectrum at low redshift, but also in the power spectra of realistically selected galaxy samples at $0 < z < 3$. Present galaxy surveys are marginally able to detect the baryonic features at low redshifts[42,43]. If future surveys improve on this and reach sufficient volume and galaxy density also at high redshift, then precision measurements of galaxy clustering will shed light on one of the most puzzling components of the Universe: the elusive dark energy field.

METHODS

The Millennium Simulation was carried out with a customized version of the GADGET2 (ref. 44) code, using the 'TreePM' method[45] for evaluating gravitational forces. This is a combination of a hierarchical multipole expansion, or 'tree' algorithm[46], and a classical Fourier-transform particle-mesh method[47]. The calculation was performed on 512 processors of an IBM p690 parallel computer at the Computing Centre of the Max-Planck Society in Garching, Germany. It used almost all of the 1 terabyte of physically distributed memory available. It required about 350,000 processor hours of CPU time, or 28 days of wall-clock time. The mean sustained floating-point performance (as measured by hardware counters) was about 0.2 teraflops, so the total number of floating-point operations carried out was of the order of 5×10^{17}.

Parameters and initial conditions. The cosmological parameters of our ΛCDM-simulation are: $\Omega_{\rm m} = \Omega_{\rm dm} + \Omega_{\rm b} = 0.25$, $\Omega_{\rm b} = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$, and $\sigma_8 = 0.9$. Here $\Omega_{\rm m}$ denotes the total matter density in units of the critical density for closure, $\rho_{\rm crit} = 3H_0^2/(8\pi G)$, where G is Newton's gravitational constant. Similarly, $\Omega_{\rm b}$ and Ω_Λ denote the densities of baryons and dark energy at the present day. The Hubble constant is parameterized as $H_0 = 100h\,{\rm km\,s^{-1}\,Mpc^{-1}}$, while σ_8 is the root-mean-square (r.m.s.) linear mass fluctuation within a sphere of radius $8h^{-1}$ Mpc extrapolated to $z = 0$. Our adopted parameter values are consistent with a combined analysis of the 2dFGRS (ref. 48) and first-year WMAP data[2].

The simulation volume is a periodic box of size $500h^{-1}$ Mpc and individual particles have a mass of $8.6 \times 10^8 h^{-1} M_\odot$. This volume is large enough to include interesting rare objects, but still small enough that the haloes of all luminous galaxies brighter than $0.1L_*$ (where L_* is the characteristic luminosity of galaxies)

are resolved with at least 100 particles. At present, the richest clusters of galaxies contain about three million particles. The gravitational force law is softened isotropically on a co-moving scale of $5h^{-1}$ kpc (Plummer-equivalent), which may be taken as the spatial resolution limit of the calculation. Thus, our simulation achieves a dynamic range of 10^5 in three dimensions, and this resolution is available everywhere in the simulation volume.

Initial conditions were laid down by perturbing a homogeneous, 'glass-like', particle distribution[49] with a realization of a gaussian random field with the ΛCDM linear power spectrum as given by the Boltzmann code CMBFAST[50]. The displacement field in Fourier space was constructed using the Zel'dovich approximation, with the amplitude of each random phase mode drawn from a Rayleigh distribution. The simulation started at redshift $z = 127$ and was evolved to the present using a leapfrog integration scheme with individual and adaptive timesteps, with up to 11,000 timesteps for individual particles. We stored the full particle data at 64 output times, each of size 300 gigabytes, giving a raw data volume of nearly 20 terabytes. This allowed the construction of finely resolved hierarchical merging trees for tens of millions of haloes and for the subhaloes that survive within them. A galaxy catalogue for the full simulation, typically containing $\sim 2 \times 10^7$ galaxies at $z = 0$ together with their full histories, can then be built for any desired semi-analytic model in a few hours on a small workstation cluster.

Galaxy-formation modelling. The semi-analytic model itself can be viewed as a simplified simulation of the galaxy-formation process, where the star formation and its regulation by feedback processes is parameterized in terms of simple analytic physical models. These models take the form of differential equations for the time evolution of the galaxies that populate each hierarchical merging tree. In brief, these equations describe radiative cooling of gas, star formation, growth of supermassive black holes, feedback processes by supernovae and AGN, and effects due to a re-ionising ultraviolet background. The morphological transformation of galaxies and the process of metal enrichment are modelled as well. To make direct contact with observational data, we apply modern population synthesis models to predict spectra and magnitudes for the stellar light emitted by galaxies, also including simplified models for dust obscuration. In this way we can match the passbands commonly used in observations.

The basic elements of galaxy-formation modelling follow previous studies[16,18–23] (see also Supplementary Information), but we also use new approaches in a number of areas. First, and of substantial importance, is our tracking of dark matter substructure. This we carry out consistently and with unprecedented resolution throughout our large cosmological volume, allowing an accurate determination of the orbits of galaxies within larger structures, as well as robust estimates of the survival time of structures infalling into larger objects. Also, we use dark matter halo properties, such as angular momentum or density profile, to determine directly sizes of galactic disks and their rotation curves. Second, we use a new model for the build-up of a population of supermassive black holes in the Universe. To this end we extend the quasar model developed in previous work[17] with a 'radio mode', which describes the feedback activity of central AGN in groups and clusters of galaxies. Although largely unimportant for the cumulative growth of the total black-hole mass density in the Universe, our results show that the radio mode becomes important at low redshift, where it has a strong impact on cluster cooling flows, reddening and reducing the brightness of central cluster galaxies. This shapes the bright end of the galaxy luminosity function, bringing our predictions into good agreement with observation.

Received 4 January; accepted 25 March 2005.

1. Bennett, C. L. *et al.* First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. *Astrophys. J. Suppl.* **148,** 1–27 (2003).
2. Spergel, D. N. *et al.* First-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Determination of cosmological parameters. *Astrophys. J. Suppl.* **148,** 175–194 (2003).
3. Riess, A. G. *et al.* Observational evidence from supernovae for an accelerating universe and a cosmological constant. *Astron. J.* **116,** 1009–1038 (1998).
4. Perlmutter, S. *et al.* Measurements of omega and lambda from 42 high-redshift supernovae. *Astrophys. J.* **517,** 565–586 (1999).
5. White, S. D. M., Navarro, J. F., Evrard, A. E. & Frenk, C. S. The baryon content of galaxy clusters: a challenge to cosmological orthodoxy. *Nature* **366,** 429–433 (1993).
6. Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.* **292,** 371–394 (1985).
7. Colberg, J. M. *et al.* Clustering of galaxy clusters in cold dark matter universes. *Mon. Not. R. Astron. Soc.* **319,** 209–214 (2000).
8. Evrard, A. E. *et al.* Galaxy clusters in Hubble volume simulations: Cosmological constraints from sky survey populations. *Astrophys. J.* **573,** 7–36 (2002).
9. Wambsganss, J., Bode, P. & Ostriker, J. P. Giant arc statistics in concord with a concordance lambda cold dark matter universe. *Astrophys. J.* **606,** L93–L96 (2004).
10. Bond, J. R., Kofman, L. & Pogosyan, D. How filaments of galaxies are woven into the cosmic web. *Nature* **380,** 603–606 (1996).
11. Jenkins, A. *et al.* The mass function of dark matter haloes. *Mon. Not. R. Astron. Soc.* **321,** 372–384 (2001).
12. Reed, D. *et al.* Evolution of the mass function of dark matter haloes. *Mon. Not. R. Astron. Soc.* **346,** 565–572 (2003).
13. Sheth, R. K. & Tormen, G. An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *Mon. Not. R. Astron. Soc.* **329,** 61–75 (2002).
14. Press, W. H. & Schechter, P. Formation of galaxies and clusters of galaxies by self-similar gravitational condensation. *Astrophys. J.* **187,** 425–438 (1974).
15. Efstathiou, G. & Rees, M. J. High-redshift quasars in the Cold Dark Matter cosmogony. *Mon. Not. R. Astron. Soc.* **230,** 5–11 (1988).
16. Springel, V., White, S. D. M., Tormen, G. & Kauffmann, G. Populating a cluster of galaxies. – I. Results at z = 0. *Mon. Not. R. Astron. Soc.* **328,** 726–750 (2001).
17. Kauffmann, G. & Haehnelt, M. A unified model for the evolution of galaxies and quasars. *Mon. Not. R. Astron. Soc.* **311,** 576–588 (2000).
18. White, S. D. M. & Frenk, C. S. Galaxy formation through hierarchical clustering. *Astrophys. J.* **379,** 52–79 (1991).
19. Kauffmann, G., White, S. D. M. & Guiderdoni, B. The formation and evolution of galaxies within merging dark matter haloes. *Mon. Not. R. Astron. Soc.* **264,** 201–218 (1993).
20. Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F. & Zepf, S. E. A recipe for galaxy formation. *Mon. Not. R. Astron. Soc.* **271,** 781–806 (1994).
21. Baugh, C. M., Cole, S. & Frenk, C. S. Evolution of the Hubble sequence in hierarchical models for galaxy formation. *Mon. Not. R. Astron. Soc.* **283,** 1361–1378 (1996).
22. Somerville, R. S. & Primack, J. R. Semi-analytic modelling of galaxy formation: the local Universe. *Mon. Not. R. Astron. Soc.* **310,** 1087–1110 (1999).
23. Kauffmann, G., Colberg, J. M., Diaferio, A. & White, S. D. M. Clustering of galaxies in a hierarchical universe. – I. Methods and results at z = 0. *Mon. Not. R. Astron. Soc.* **303,** 188–206 (1999).
24. Fan, X. *et al.* A survey of z > 5.7 quasars in the Sloan Digital Sky Survey. II. Discovery of three additional quasars at z > 6. *Astron. J.* **125,** 1649–1659 (2003).
25. Fan, X. *et al.* A survey of z > 5.7 quasars in the Sloan Digital Sky Survey. III. Discovery of five additional quasars. *Astron. J.* **128,** 515–522 (2004).
26. Tremaine, S. *et al.* The slope of the black hole mass versus velocity dispersion correlation. *Astrophys. J.* **574,** 740–753 (2002).
27. Merritt, D. & Ferrarese, L. Black hole demographics from the M_{BH}-σ relation. *Mon. Not. R. Astron. Soc.* **320,** L30–L34 (2001).
28. Hawkins, E. *et al.* The 2dF Galaxy Redshift Survey: correlation functions, peculiar velocities and the matter density of the universe. *Mon. Not. R. Astron. Soc.* **346,** 78–96 (2003).
29. Benson, A. J., Cole, S., Frenk, C. S., Baugh, C. M. & Lacey, C. G. The nature of galaxy bias and clustering. *Mon. Not. R. Astron. Soc.* **311,** 793–808 (2000).
30. Weinberg, D. H., Davé, R., Katz, N. & Hernquist, L. Galaxy clustering and galaxy bias in a ΛCDM universe. *Astrophys. J.* **601,** 1–21 (2004).
31. Padilla, N. D. & Baugh, C. M. The power spectrum of galaxy clustering in the APM survey. *Mon. Not. R. Astron. Soc.* **343,** 796–812 (2003).
32. Zehavi, I. *et al.* On departures from a power law in the galaxy correlation function. *Astrophys. J.* **608,** 16–24 (2004).
33. Norberg, P. *et al.* The 2dF Galaxy Redshift Survey: luminosity dependence of galaxy clustering. *Mon. Not. R. Astron. Soc.* **328,** 64–70 (2001).
34. Zehavi, I. *et al.* Galaxy clustering in early Sloan Digital Sky Survey redshift data. *Astrophys. J.* **571,** 172–190 (2002).
35. Madgwick, D. S. *et al.* The 2dF Galaxy Redshift Survey: galaxy clustering per spectral type. *Mon. Not. R. Astron. Soc.* **344,** 847–856 (2003).
36. de Bernardis, P. *et al.* A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature* **404,** 955–959 (2000).
37. Mauskopf, P. D. *et al.* Measurement of a peak in the Cosmic Microwave Background power spectrum from the North American test flight of Boomerang. *Astrophys. J.* **536,** L59–L62 (2000).
38. Blake, C. & Glazebrook, K. Probing dark energy using baryonic oscillations in the galaxy power spectrum as a cosmological ruler. *Astrophys. J.* **594,** 665–673 (2003).
39. Jenkins, A. *et al.* Evolution of structure in cold dark matter universes. *Astrophys. J.* **499,** 20–40 (1998).
40. Bardeen, J. M., Bond, J. R., Kaiser, N. & Szalay, A. S. The statistics of peaks of Gaussian random fields. *Astrophys. J.* **304,** 15–61 (1986).
41. Adelberger, K. L. *et al.* A counts-in-cells analysis of Lyman-break galaxies at redshift Z = 3. *Astrophys. J.* **505,** 18–24 (1998).
42. Cole, S. *et al.* The 2dF Galaxy Redshift Survey: Power-spectrum analysis of the final dataset and cosmological implications. *Mon. Not. R. Astron. Soc.* (submitted); preprint at ⟨http://xxx.lanl.gov/astro-ph/0501174⟩ (2005).
43. Eisenstein, D. J. *et al.* Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies. *Astrophys. J.* (submitted);

preprint at ⟨http://xxx.lanl.gov/astro-ph/0501171⟩ (2005).
44. Springel, V., Yoshida, N. & White, S. D. M. GADGET: a code for collisionless and gasdynamical cosmological simulations. *N. Astron.* **6,** 79–117 (2001).
45. Xu, G. A new parallel n-body gravity solver: TPM. *Astrophys. J. Suppl.* **98,** 355–366 (1995).
46. Barnes, J. & Hut, P. A hierarchical *O*(*N* log*N*) force-calculation algorithm. *Nature* **324,** 446–449 (1986).
47. Hockney, R. W. & Eastwood, J. W. *Computer Simulation Using Particles* Ch. 5 (McGraw-Hill, New York, 1981).
48. Colless, M. *et al.* The 2dF Galaxy Redshift Survey: spectra and redshifts. *Mon. Not. R. Astron. Soc.* **328,** 1039–1063 (2001).
49. White, S. D. M. in *Cosmology and Large-Scale Structure* (eds Schaefer, R., Silk, J., Spiro, M. & Zinn-Justin, J.) Ch. 8 (Elsevier, Dordrecht, 1996).
50. Seljak, U. & Zaldarriaga, M. A line-of-sight integration approach to Cosmic Microwave Background anisotropies. *Astrophys. J.* **469,** 437–444 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The computations reported here were performed at the Rechenzentrum der Max-Planck-Gesellschaft in Garching, Germany.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to V.S. (vspringel@mpa-garching.mpg.de).