

# On the Reliability of Initial Conditions for Dissipationless Cosmological Simulations

Alexander Knebe<sup>1</sup> and Alvaro Domínguez<sup>2</sup>

<sup>1</sup> Centre for Astrophysics and Supercomputing, Swinburne University, PO Box 218, Mail #31,  
Hawthorn, Victoria 3122, Australia  
aknebe@astron.swin.edu.au

<sup>2</sup> Max-Planck-Institute für Metallforschung, Heisenbergstrasse 3, 70569 Stuttgart, Germany

Received 2002 October 2, accepted 2003 February 21

**Abstract:** We present the study of ten random realisations of a density field characterised by a cosmological power spectrum  $P(k)$  at redshift  $z = 50$ . The reliability of such initial conditions for  $N$ -body simulations is tested with respect to their correlation properties. The power spectrum  $P(k)$  and the mass variance  $\sigma_M(r)$  do not show detectable deviations from the desired behaviour in the intermediate range of scales between the mean interparticle distance and the simulation volume. The estimator for  $\xi(r)$  is too noisy to detect any reliable signal at the initial redshift  $z = 50$ . The particle distributions are then evolved forward until  $z = 0$ . This allows us to explore the cosmic variance stemming from the random nature of the initial conditions. With cosmic variance we mean the fact that a simulation represents a single realisation of the stochastic initial conditions whereas the real Universe contains many realisations of regions of the size of the box; this problem affects most importantly the scales at about the fundamental mode. We study morphological descriptors of the matter distribution such as the genus, as well as the internal properties of the largest object(s) forming in the box. We find that the scatter is at least comparable to the scatter in the fundamental mode.

**Keywords:** methods:  $N$ -body simulations — dark matter

## 1 Introduction

Our present understanding of the formation and properties of the cosmological large-scale structure relies to a large extent on  $N$ -body simulations: given the difficulty in addressing theoretically the highly nonlinear regime of the growth of density inhomogeneities by the gravitational instability, simulations have proven a valuable tool to get insight into the (nonlinear) structure formation scenarios. Therefore, it is of considerable importance to confirm the reliability of such simulations.

It has been claimed recently (Baertschiger & Sylos Labini 2002) that there are major problems with generating initial conditions (ICs) for  $N$ -body simulations. We can identify several reasons why the ICs may introduce uncertainties in the subsequent evolution. First, there is the problem of finite-mass resolution or *discreteness*: the initial continuum density field is modelled by the distribution of a *finite* number of point particles  $N$ , therefore only a finite number of Fourier modes of the density field can be reproduced reliably. The maximum wavenumber (*Nyquist wavenumber*) is given by  $k_{\max} = \pi/\Delta x$ , where  $\Delta x$  is the mean interparticle separation. The modes  $k \gtrsim k_{\max}$  have spurious values related to the point-particle distribution and may lead to artificial effects in the posterior dynamical evolution. The finite-mass resolution is expected to be irrelevant if the nonlinear mode–mode coupling to the modes  $k \gtrsim k_{\max}$  has only a small influence on the dynamics.

The second problem with the ICs is the finite size of the simulation box with side length  $B$ , which implies that the values of the Fourier modes of the density field with

wavenumber smaller than the *fundamental wavenumber*,  $k < k_{\min} = 2\pi/B$ , are artificially set to zero. This leads to two possible difficulties: first, the absence of mode–mode coupling to those large-scale modes, and second the so-called *cosmic variance*, meaning that the simulation box represents only one (finite-sized) realisation of the stochastic initial density field, whereas the true Universe contains many realisations of regions of the size of the box. Therefore, the morphological properties of the matter distribution in a certain volume, as measured by, for example, the genus statistics, will presumably show some intrinsic scatter when placing the volume at different locations in the real Universe. And this will also happen with the (internal) properties of any given class of objects. This is one of the main aspects of the current study and what we refer to as cosmic variance (in  $N$ -body simulations) throughout the paper even though one might argue that this is not the ‘real’ cosmic variance but rather an artificially introduced *sampling variance*. However, we are actually interested in exactly that (sampling) effect which can easily be tested by simply running the same cosmological simulation but using different random realisations of the initial density field.

In this work we study systematically the reliability of the initial density field used as an input to the  $N$ -body simulations as well as the effect of their random nature onto the internal properties of clusters. In Section 2 we briefly review the most commonly used way to generate ICs and the code used to evolve the particles into the nonlinear regime. Section 3 focuses on some of the statistical characteristics of the dark matter field: we consider

the 2-point correlation function, the power spectrum, the mass variance in spheres, and the Minkowski functionals, the latter being sensitive to correlations of higher order. Finally, in Section 4 we investigate dark matter clusters identified within the simulations and quantify the effect of the cosmic variance on their internal properties.

## 2 N-body Simulations

### 2.1 Generating Initial Conditions

The commonly used way for setting up initial conditions for a cosmological simulation is to make use of the Zeldovich approximation to move particles from a Lagrangian point  $\vec{q}$  to a Eulerian point  $\vec{x}$  (e.g. Efstathiou, Frenk, & White 1985):

$$\vec{x} = \vec{q} - D(t)\vec{S}(\vec{q}), \quad (1)$$

where  $D(t)$  describes the growing mode of linear fluctuations and  $\vec{S}(\vec{q})$  is the ‘displacement field’. This method is neither restricted to a cosmological scenario nor to the Zeldovich approximation: it is very general, relying only on the continuity equation for the transport of particles in the limit  $D(t) \rightarrow 0$ . The initial Lagrangian coordinates  $\vec{q}$  are usually chosen to form a regular, three-dimensional lattice although there are other possible point-particle realisations yielding a homogeneous and isotropic density field on large scales (i.e. glass-like initial conditions, White 1996).

For the runs presented in this study we used the code described in Klypin & Holtzman (1997) to set up the initial conditions

$$\begin{aligned} \vec{S}(\vec{q}) &= \nabla_q \Phi(\vec{q}), \\ \Phi(\vec{q}) &= \sum_{\vec{k}} a_{\vec{k}} \cos(\vec{k} \cdot \vec{q}) + b_{\vec{k}} \sin(\vec{k} \cdot \vec{q}), \end{aligned} \quad (2)$$

where the Fourier coefficients  $a_{\vec{k}}$  and  $b_{\vec{k}}$  are related to a pre-calculated input power spectrum of density fluctuations,  $P(k)$ , as follows:

$$a_{\vec{k}} = R_1 \frac{1}{k^2} \sqrt{P(k)}, \quad b_{\vec{k}} = R_2 \frac{1}{k^2} \sqrt{P(k)}. \quad (3)$$

$R_1, R_2$  are (Gaussian) random numbers with mean zero and dispersion unity. The factor  $1/k^2$  is (the Fourier transform of) the Green’s function of Poisson’s equation<sup>1</sup> and  $\Phi(\vec{q})$  can therefore be understood as the gravitational potential created by a Gaussian stochastic density field whose power spectrum agrees with the input  $P(k)$ ; the power spectrum  $P(k)$  measures the strength of each individual  $k$ -mode contributing to the density field. However, to fully preserve the random nature both amplitudes (sine- and cosine-wave) are to be picked from a Gaussian distribution.

<sup>1</sup>Actually,  $-1/k^2$  is the correct Green’s function, but the factor  $-1$  can be dropped as  $R_1$  and  $R_2$  scatter around zero.

Equation (2) can be rewritten introducing complex numbers:

$$\begin{aligned} \Phi(\vec{q}) &= \sum_{\vec{k}} A_{\vec{k}} \exp(i[\vec{k} \cdot \vec{q} + \theta_{\vec{k}}]), \\ A_{\vec{k}} \exp(i\theta_{\vec{k}}) &:= \frac{1}{2}[a_{\vec{k}} + a_{-\vec{k}} - i(b_{\vec{k}} - b_{-\vec{k}})]. \end{aligned} \quad (4)$$

Both  $A_{\vec{k}}$  and  $\theta_{\vec{k}}$  need to be drawn from appropriate random distributions. However, the ICs of cosmological relevance are ergodic for  $A_{\vec{k}}$  with  $k \gg k_{\min}$ , making their random nature irrelevant: spatial regions of size much smaller than the simulation box already work as many different realisations inside the box for those amplitudes. Thus, cosmic variance enters through the random nature of the phases  $\theta_{\vec{k}}$  and of the amplitudes  $A_{\vec{k}}$  for  $k \approx k_{\min}$ .

The idea of this paper is to create a certain number of random realisations of the same power spectrum  $P(k)$  by using different random seeds when drawing  $R_1, R_2$  in equation (3). The input power spectrum  $P(k)$  was calculated using the CMBFAST code (Seljak & Zaldarriaga 1996), and all parameters (e.g. box size, number of particles, force resolution, integration steps) were fixed except for the seed for generating the random sequence providing the  $R$ -values.<sup>2</sup>

### 2.2 Simulation Details

We created a database of ten simulations that all were started at a redshift  $z_i = 50$  and evolved until  $z = 0$  in a  $\Lambda$ CDM ( $\Omega_0 = 0.3$ ,  $\Omega_\lambda = 0.7$ ,  $\Omega_b h^2 = 0.04$ ,  $h = 0.7$ ,  $\sigma_8 = 0.9$ ) cosmological model using  $128^3$  particles within a box of side length  $B = 64 h^{-1}$  Mpc, giving a mass resolution of  $m_p = 1.04 \times 10^{10} h^{-1} M_\odot$ . They were carried out using the publicly available adaptive mesh refinement code MLAPM (Knebe, Green, & Binney 2001). MLAPM reaches high force resolution by refining all high-density regions with an automated refinement algorithm. The refinements are recursive: the refined regions can also be refined, each subsequent refinement having cells that are half the size of the cells in the previous level. This creates an hierarchy of refinement meshes of different resolutions covering regions of interest. The refinement is done cell by cell (individual cells can be refined or de-refined) and meshes are not constrained to have a rectangular (or any other) shape. The criterion for (de-)refining a cell is simply the number of particles within that cell and a detailed study of the appropriate choice for this number can be found elsewhere (Knebe et al. 2001). The code also uses multiple time steps on different refinement levels where the time step for each refinement level is two times smaller than the step on the previous level. A regular  $256^3$  domain grid was used to cover the whole computational volume in all runs, and cells were refined as soon as the number of particles per cell exceeded the preselected value of 8. We stored snapshots of the particle distribution at redshifts  $z = 5$ ,  $z = 1$ ,  $z = 0.5$ ,  $z = 0.25$ , and  $z = 0$ . At the end of the runs the force resolution is determined by the

<sup>2</sup>An appropriate routine might be GASDEV from Press et al. (1992).

highest refinement level reached: for the runs at hand the finest grid at  $z = 0$  consisted of 8192 cells per side and was called into existence at redshift  $z \sim 0.88$ . This grid corresponds to a force resolution of about  $23 h^{-1}$  kpc which is simply three times the grid spacing and gives the scale where the forces are purely Newtonian. This is sufficient for this study as we are mainly interested in the overall (large-scale) clustering properties. But as we will see later on, we are resolving approximately 2% of the virial radius of the most massive halo formed in the runs, which is sufficient for investigations of the internal properties such as velocity dispersion, spin parameter, and triaxiality.

### 3 Analysis I: Dark Matter Field

We first focus on the properties of the dark matter particle distributions. Our main aim is to assess the recent claims by Baertschiger & Sylos Labini (2002) that there are major problems with generating initial conditions for  $N$ -body simulations in the way as outlined in Section 2.1. Because it is common to use either a regular grid or a glass-like distribution as Lagrangian starting points  $\vec{q}$  for the Zeldovich approximation (cf. equation(1)), their arguments try to prove that this leads to spurious artifacts related to, for example, the regular structure of such a grid, and that the initial conditions are not able to reflect the superposed CDM-like fluctuations at all.

#### 3.1 Power Spectra

When creating a fluctuating density field in a certain volume by using a fixed number of particles, one is limited in the range of  $ks$  by the size of that volume on the one hand, and the number of particles used to sample the waves on the other hand. The wavenumber of the lowest frequency wave (fundamental mode) to fit into the box is given by  $k_{\min} = 2\pi/B$  where  $B$  is the side length of the box. The maximum wavenumber is determined by the Nyquist frequency,  $k_{\max} = \pi/\Delta x$ , where  $\Delta x = B/N^{1/3}$  is the mean particle separation (not to be confused with the grid spacing used in the  $N$ -body code and for extracting the power spectrum from such a particle distribution, respectively). A recent investigation showed that high-resolution  $N$ -body simulations where even smaller scales than  $k_{\max}$  are resolved are justified for power spectra with an effective spectral index  $n_{\text{eff}} = d \log P(k)/d \log k$  much less than  $-1$  (Hamana, Yoshida, & Suto 2001). And this is the case for (nearly) all CDM type spectra as  $P(k) \propto k^{-3}$  for large  $k$ . The evolution of power on small scales is driven by the transfer of power from large scales and hence it is important to follow that evolution with an appropriate force resolution even though that small-scale power was not present in the initial conditions (see Introduction).

Using the particle data at redshifts  $z = 50$  (initial conditions),  $z = 5$ ,  $z = 1$ , and  $z = 0$ , we derived  $P(k)$  by Fourier transforming the density field on a regular  $512^3$  grid,<sup>3</sup> which effectively introduces  $k_{\max} \approx 25 h \text{ Mpc}^{-1}$  as the

maximum wavenumber to be recovered from the data. We adopted the method for extracting even higher  $k$  waves from the particle distributions as outlined in Jenkins et al. (1998). The power spectra were then compared to the nonlinear prediction given by Peacock & Dodds (1996, hereafter PD96). However, their fitting parameters depend on the spectral slope  $n = d \ln P(k)/d \ln k$  and hence some recipe needs to be adopted when applying it to a cosmological  $P(k)$  where  $n$  is a function of  $k$ . We used  $n_{\text{eff}}$  defined via

$$n_{\text{eff}}(k_l) = \left. \frac{d \ln P(k)}{d \ln k} \right|_{k=k_l/2} \quad (5)$$

for the estimate of the spectral index  $n_{\text{eff}}$  at wavenumber  $k_l$  (cf. Jenkins et al. 1998; PD96).

In Figure 1 the results are shown along with the linearly extrapolated  $P(k)$ . There are a couple of things to note besides the overall good agreement of the estimated  $P(k)$  with the PD96 prediction: first, the power spectrum derived from the particle distribution agrees extremely well with the input  $P(k)$ <sup>4</sup> and the fundamental mode has not turned nonlinear at  $z = 0$ ; second, we can clearly see how the scatter prominent in the large waves  $k_{\min} \gtrsim 2\pi/B$  at high redshifts moves towards higher  $k$  values at later times. The scatter in the initial conditions is of the order of 20% and it arises because only a small number of such harmonics do fit into the finite box of side length  $B$ . We ascribe the migration of the scatter downwards to smaller scales (higher  $k$  values) to the transfer of power from large to small scales: the higher the amplitude  $A_{\vec{k}}$  at  $k_{\min}$  (cf. equation(4)) the more power can be transferred to smaller  $ks$ . And hence we are facing a faster evolution of small-scale structures leading to the observed dispersion amongst the individual runs.

The discrepancy with the PD96 prediction for  $z = 5$  for  $k > 10$  is not physical. Even though the MLAPM code already invoked three levels of refinements at  $z = 5$  they are still very small in size, i.e. there are only approximately 40 000 refinement cells in total with about 12 000 particles ( $\sim 0.6\%$  of all particles) being moved on those levels. However, a visual inspection of the refined regions shows that the grids are covering all prospective halo formation sites and hence we are following the build-up of structures correctly. But when trying to recover those high- $k$  modes from the simulations we are left with an observed mismatch due to the majority of the particles still being moved on the  $256^3$  grid.

#### 3.2 Mass Variance $\sigma(r)$

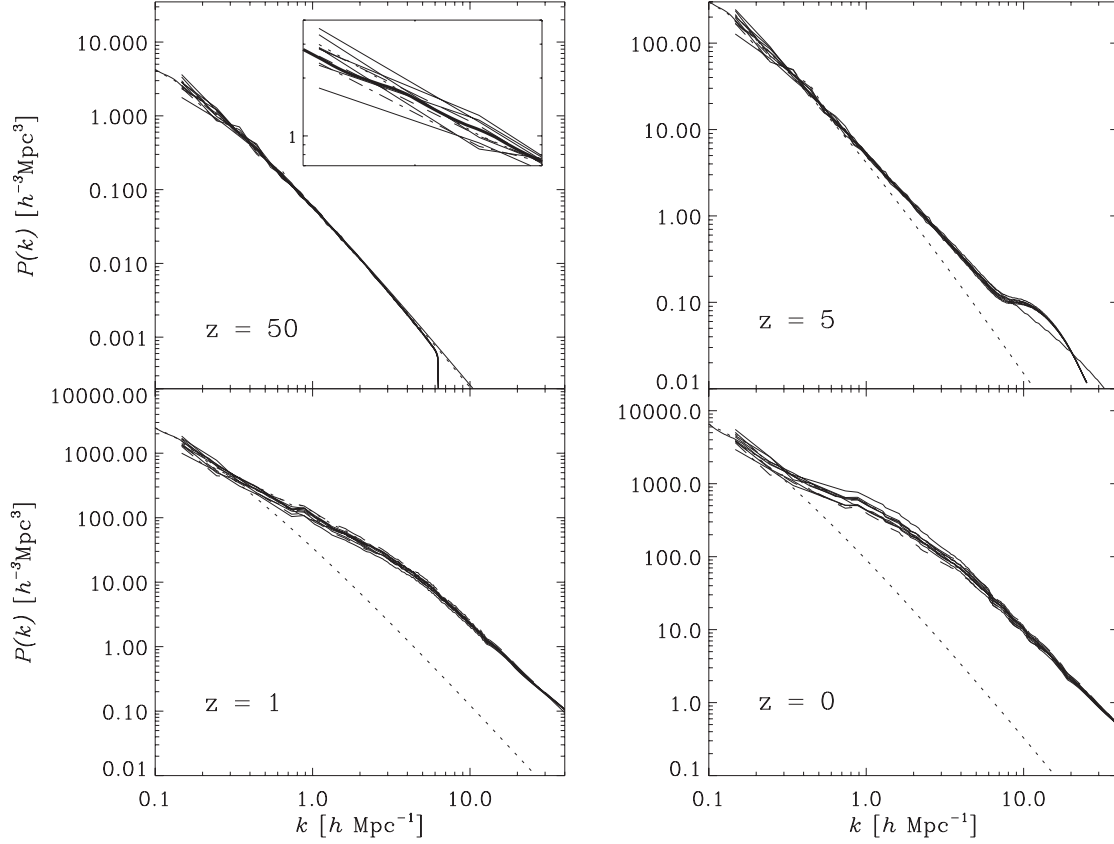
The variance  $\sigma_M$  of the mass in spheres with radius  $r$  is given by

$$\sigma_M^2(r) = \frac{1}{2\pi^2} \int_0^{+\infty} P(k) \hat{W}^2(kr) k^2 dk, \quad (6)$$

$$\hat{W}(x) = \frac{3}{x^3} (\sin x - x \cos x).$$

<sup>3</sup>The density was assigned to the grid cells using the triangular shape cloud method.

<sup>4</sup>At redshift  $z = 50$  the nonlinear  $P(k)$  as given by PD96 is indistinguishable from the linear  $P(k)$ .



**Figure 1** Power spectrum evolution for all ten runs as compared to the prediction by Peacock & Dodds (1996) (—) and the linear  $P(k)$  (---), respectively. The inset panel for  $z = 50$  focuses on the fundamental mode  $k_{\min} = 2\pi/B$  which shows a  $1\sigma$  variance of approximately 20%.

The function  $\sigma_M^2(r)$  is readily calculated and can be compared to an adequate estimator  $\sigma_{M,\text{est}}^2(r)$  when being applied to the actual particle data. Our estimator distributes a certain number of spheres with radius  $r$  at random in the simulation volume and compares the number of particles inside those spheres to the expected mean value

$$\sigma_{M,\text{est}}^2(r) = \frac{1}{\langle N_r \rangle^2} \sum_{i=1}^{N_s} \frac{(N_i(r) - \langle N_r \rangle)^2}{N_s - 1}. \quad (7)$$

$N_s$  is the total number of spheres with radius  $r$  and  $\langle N_r \rangle = \langle \rho \rangle 4\pi r^3 / 3m_p$  is the mean number of particles in such a sphere.

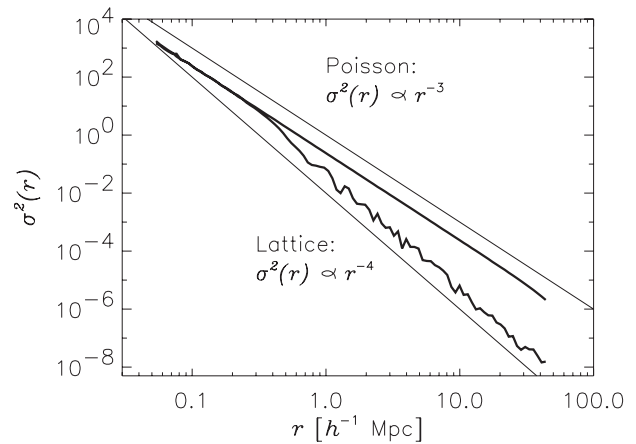
### 3.2.1 Reliability of Estimator

In order to make sure our estimator works as expected we started by applying it to particle distributions where simple scaling laws for  $\sigma_M^2(r)$  can be calculated analytically. For a purely Poissonian particle distribution one easily derives

$$\sigma_{M,\text{Poisson}}^2(r) \propto r^{-3}, \quad (8)$$

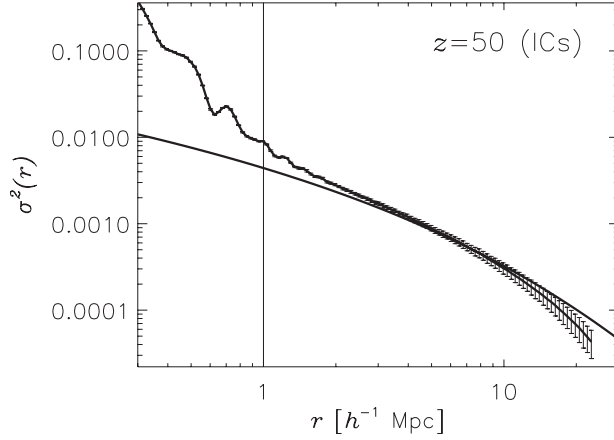
and for a ‘shuffled’ lattice (e.g. Gabrielli, Joyce, & Sylos Labini 2002)

$$\sigma_{M,\text{Lattice}}^2(r) \propto r^{-4}, \quad (r \gg \text{lattice spacing}). \quad (9)$$



**Figure 2** Reliability check for our  $\sigma_{M,\text{est}}^2(r)$  estimator equation (7). The solid lines have the slopes of the analytical expectations (refer to the text for further details). All amplitudes are arbitrary.

From Figure 2 we deduce that our estimator does indeed work correctly: we created ten Poisson distributions of  $128^3$  particles in a  $(128 h^{-1} \text{ Mpc})^3$  volume and for each distribution we calculated  $\sigma_{M,\text{est}}^2(r)$  using 10 000 spheres (for each  $r$  value). The curve shown is the average taken over the ten Poisson distributions. The error bars are too small to be presented. The shuffled lattice distribution was created as follows: we placed  $128^3$  particles on the nodes of a  $128^3$  grid with spacing  $1 h^{-1} \text{ Mpc}$ , and



**Figure 3** Mean value of  $\sigma_M^2(r)$  as given by equation (7) when averaged over the ten initial conditions at redshift  $z=50$ . The error bars are  $1\sigma$  errors. The solid line is the analytical expectation given by equation (6). The vertical line indicates the scale corresponding to the (particle) Nyquist frequency,  $k_{\max}$ .

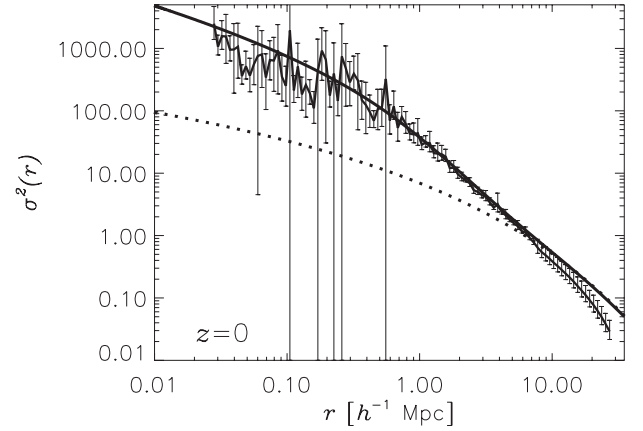
then each particle was shifted in  $x$ ,  $y$ , and  $z$  directions by a random amount uniformly distributed in the range  $[-0.05, 0.05] h^{-1} \text{ Mpc}$ ; ten such realisations were created. The curve shown in Figure 2 is again the mean estimate when averaging over the ten realisations. In both tests we recover the expected scaling relation.

### 3.2.2 Application to N-body Data

We now apply the estimator equation (7) to our ICs as well as the final outputs at redshift  $z=0$ . Figure 3 shows  $\sigma_{M,\text{est}}^2(r)$  compared to the analytical  $\sigma_M^2(r)$  as given by equation (6). For every scale  $r$  we again used  $N_s = 10\,000$  randomly placed spheres. The mean mass variance  $\langle \sigma_{M,\text{est}}^2(r) \rangle_{\text{set}}$  (averaged over the ten realisations) is plotted and the error bars are 1 times the variance  $\sigma_{M,\text{est}}^2(r)$  around the mean value  $\langle \sigma_{M,\text{est}}^2(r) \rangle_{\text{set}}$ .

Contrary to the findings of Baertschiger & Sylos Labini (2002), we observe that the initial conditions agree, from approximately the scale of the particle Nyquist frequency out to nearly half the box size, with the analytical predictions. The faster drop of  $\langle \sigma_{M,\text{est}}^2(r) \rangle_{\text{set}}$  for scales approaching the box size is simply the effect of the finite (periodical) box. As soon as the volume of the sphere comes close to the actual box size (which happens for  $r \approx B/2$ ) one finds nearly all particles in the sphere due to the periodic boundary conditions. Hence the variance  $\sigma_{M,\text{est}}^2(r)$  drops faster than predicted by equation (6). And the larger amplitude of  $\sigma_{M,\text{est}}^2(r)$  for small scales is indeed a reflection of the discreteness of the initial conditions. But in any case Figure 3 is a rather convincing argument that the mass variance in the initial conditions does agree with the CDM type fluctuations as described by the input power spectrum  $P(k)$ .

When comparing  $\sigma_{M,\text{est}}^2(r)$  for the final output at redshift  $z=0$  to the analytical  $\sigma_M^2(r)$  in Figure 4 (both using the linearly extrapolated  $P(k)$  as well as the non-linear  $P(k)$  given by PD96) we notice again a couple of things: firstly, the large scales ( $r \gtrsim B/2$ ) are still below



**Figure 4** Same as Figure 3 but this time for the final output at  $z=0$ . The two analytical curves are the linear extrapolation of equation (6) to  $z=0$  ( $\cdots$ ) and the prediction for  $\sigma^2(r)$  when using the PD96 power spectrum with equation (6) ( $\text{—}$ ).

the expectation, and secondly, there is more pronounced scatter on scales  $r < 0.6 h^{-1} \text{ Mpc}$  than found for the ICs. For  $r \gtrsim B/2$ , the explanation is again the finite periodic box. The increased value for the variance  $\sigma_{M,\text{est}}^2(r)$  for small scales  $r < 1 h^{-1} \text{ Mpc}$  (and its large scatter) is readily explained by the fact that gravitationally bound objects (and voids) are forming which introduces some sort of ‘semi-discreteness’: this gives rise to a higher variance (as well as larger scatter) on scales related to the average size of such objects, i.e.  $\sim 1 h^{-1} \text{ Mpc}$  and below.

### 3.3 Two-Point Correlation Function $\xi(r)$

The two-point correlation function is the Fourier transform of the power spectrum:

$$\xi(r) = \frac{1}{2\pi^2} \int_0^{+\infty} P(k) \frac{\sin(kr)}{kr} k^2 dk. \quad (10)$$

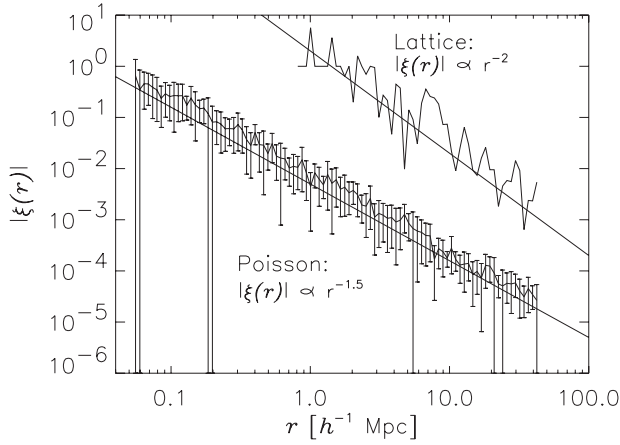
The basic interpretation of  $\xi(r)$  is that it is the average number of neighbours to a given object with distance  $r$  in excess of a Poisson distribution. And this is how we realise an estimator for  $\xi(r)$ . We start again putting down a certain number of spheres in the simulation box, but this time centred at particles. We then create a shell of thickness  $dr$  extending from  $r$  to  $r + dr$ . The correlation function can now be estimated by

$$\xi_{\text{est}}(r) = \frac{\Gamma(r, dr)}{\langle \rho \rangle} - 1, \quad (11)$$

where  $\langle \rho \rangle$  is the mean number density of the simulation and  $\Gamma(r, dr)$  the mean number density of particles found in the shell  $[r, r + dr]$ :

$$\Gamma(r, dr) = \frac{1}{N_s} \sum_{i=1}^{N_s} \Gamma_i(r, dr). \quad (12)$$

For each value of  $r$ , we again use  $N_s = 10\,000$  shells  $[r, r + dr]$  centred at a randomly chosen particle for calculating the average number density of particles  $\Gamma(r, dr)$ .



**Figure 5** Scaling relations for  $|\xi_{\text{est}}(r)|$  when applied to a set of Poisson distributions and of shuffled lattices with spacing  $1 h^{-1} \text{ Mpc}$ . The error bars for the Poisson distribution are  $1\sigma$  and for the lattice are too small to be presented. The amplitudes are again arbitrary.

### 3.3.1 Reliability of Estimator

This time it is more difficult to calibrate the estimator equation (11) because our test models (Poisson and shuffled lattice) consists of Dirac  $\delta$ s and zones of vanishing correlation.

For the Poissonian case,  $\xi(r) = 0$  if  $r \neq 0$ , so that we expect  $\xi_{\text{est}}(r)$  to fluctuate around zero with an amplitude proportional to the dispersion of the estimator,  $\langle \xi_{\text{est}}^2(r) \rangle$ . Absence of correlations allows an easy estimation of the dispersion: if  $N_s$  is not too large (so that the probability that shells overlap is small), the numbers  $\Gamma_i$  in equation (12) are uncorrelated with each other and have a Poissonian distribution. One can then show immediately

$$\langle \Gamma_i \rangle = \langle \varrho \rangle, \quad \langle \Gamma_i \Gamma_j \rangle - \langle \Gamma_i \rangle \langle \Gamma_j \rangle = \frac{\langle \Gamma_i \rangle}{4\pi r^2 dr} \delta_{ij}, \quad (13)$$

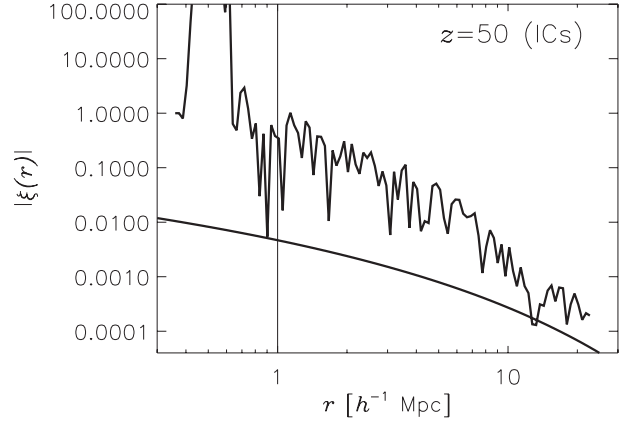
and then

$$\langle \xi_{\text{est}}^2(r) \rangle = \frac{1}{N_s \langle \varrho \rangle 4\pi r^2 dr}. \quad (14)$$

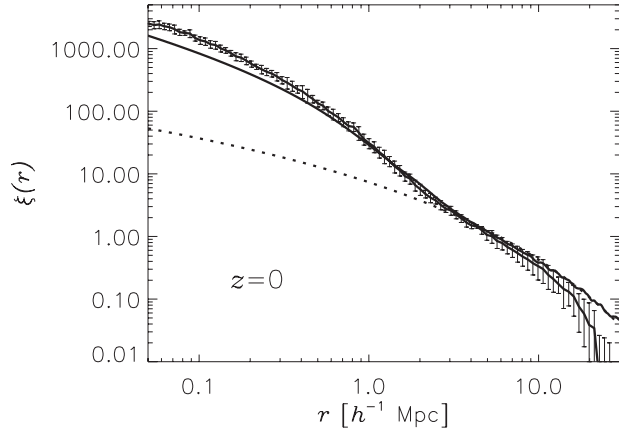
We took  $dr \propto r$  (logarithmic binning), so that we expect the amplitude of the fluctuations in  $\xi_{\text{est}}(r)$  to decay as  $r^{-3/2}$ , as indeed observed in Figure 5, where the error bars are again  $1\sigma$  errors when averaging over the ten random sets. The figure also shows  $|\xi_{\text{est}}(r)|$  for the shuffled lattice with grid spacing of  $1 h^{-1} \text{ Mpc}$ . We believe that the observed  $r^{-2}$ -decay is again due to  $\langle \xi_{\text{est}}^2(r) \rangle$ , as in the Poissonian case.

### 3.3.2 Application to N-body Data

Figure 6 shows the result of applying the estimator equation (11) to the actual  $N$ -body data at the initial redshift  $z = 50$ . We plot the absolute value  $|\xi(r)|$  as the correlation function tends to oscillate around zero, too. The curve is the average taken over the ten runs (as usual), but we do not plot the error bars as the data already show a noticeable level of noise. This noise is in fact so strong as to mask the signal (the  $\Lambda$ CDM behaviour in this case); we already



**Figure 6** Two-point correlation function for initial particle distribution at redshift  $z = 50$ . For clarity no error bars are shown due to a high level of noise. The solid line is the expected  $\xi(r)$  as given by equation (10). We plot  $|\xi(r)|$ , as the estimated function tends to oscillate around zero for scales smaller than the Nyquist wavelength (indicated by the vertical line).

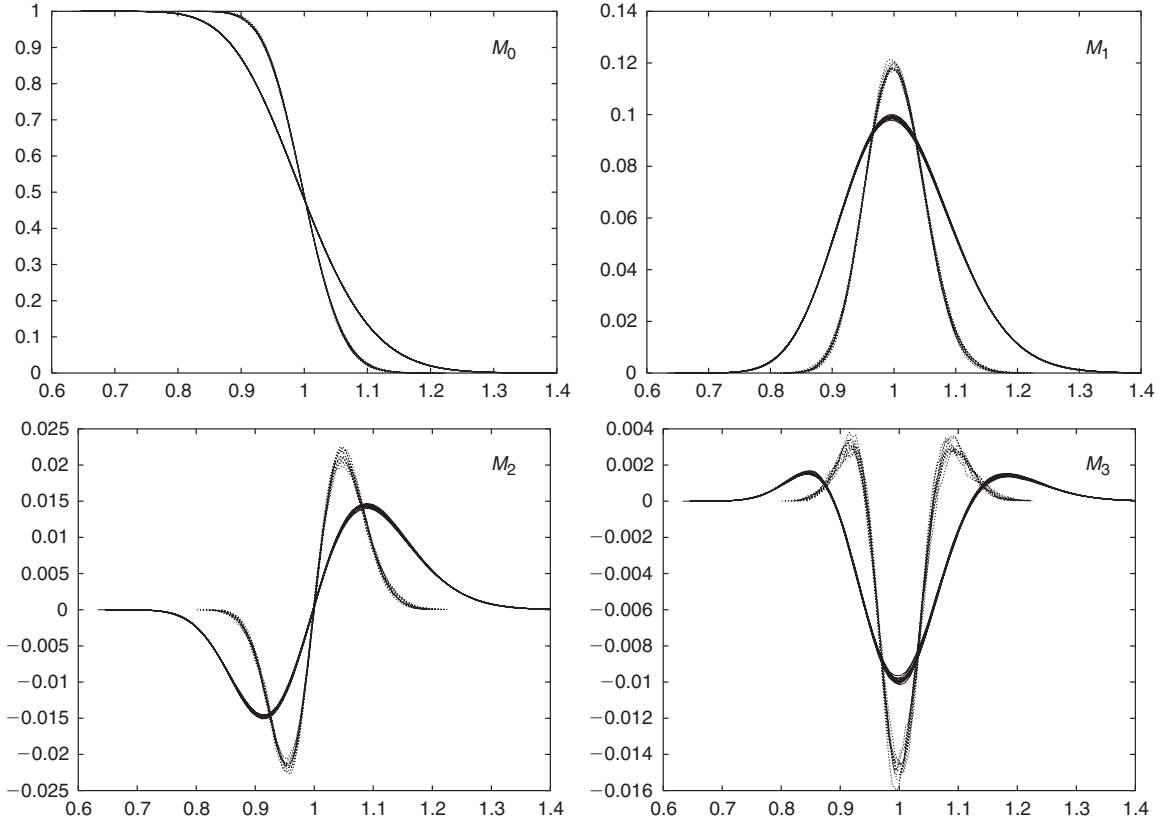


**Figure 7** Same as Figure 6 but for  $z = 0$ . The error bars are  $1\sigma$  errors when averaging over the ten runs. The two lines are again the analytical PD96 fit (—) and the linear theory (···).

found this problem with the test models especially for the lattice distribution upon which the ICs are based (cf. equation (1)). It seems that an improvement of the estimator equation (11) is required to extract reliable information in these extreme cases.

Figure 7 shows the same quantity for the  $z = 0$  data, where the analytical curves are the correlations derived from the linearly extrapolated  $P(k)$  (dotted line) and the nonlinear  $P(k)$  from PD96 (solid line) in equation (10), respectively. This time we find a deviation of the estimated  $\xi(r)$  from the one predicted using PD96 on small scales. However, the error bars are ‘only’  $1\sigma$  and the PD96 prediction still lies within the  $3\sigma$  level. One must also note that the estimator equation (11) is biased towards high density regions where most of the particles at  $z = 0$  will reside, since the shells are centred at (randomly chosen) particle positions rather than placing them randomly at any point (like the estimator equation (7) does, which also explains why the scatter for  $\xi_{\text{est}}^2(r)$  at these small scales





**Figure 8** MFs of the ten realisations of the initial matter distribution ( $z = 50$ ) as a function of the density threshold  $1 + \delta$ , at two different spatial resolutions:  $128^3$ -cell grid (—) and  $32^3$ -cell grid (···).

is much smaller than for  $\sigma_{M, \text{est}}^2(r)$  observed in Figure 4). However, we varied the number of spheres  $N_s$  from 50 to 100 000 and could only detect a mild (if any) dependence of the amplitude on  $N_s$ .

Nevertheless, if one is to believe this discrepancy, then it is not obvious which one is to be blamed, the simulations or the PD96 fit. We have confirmed the excellent agreement of PD96 with our estimated  $P(k)$  in the probed range of wavenumbers (see Figure 1). However, when using equation 10, one is extrapolating the PD96 fit to all wavenumbers. Clearly, the discrepancy should originate from the modes beyond the probed range, but with the information at hand, one cannot conclude whether their effect is estimated wrongly by the simulation or by the PD96 fit.

### 3.4 Minkowski Functionals

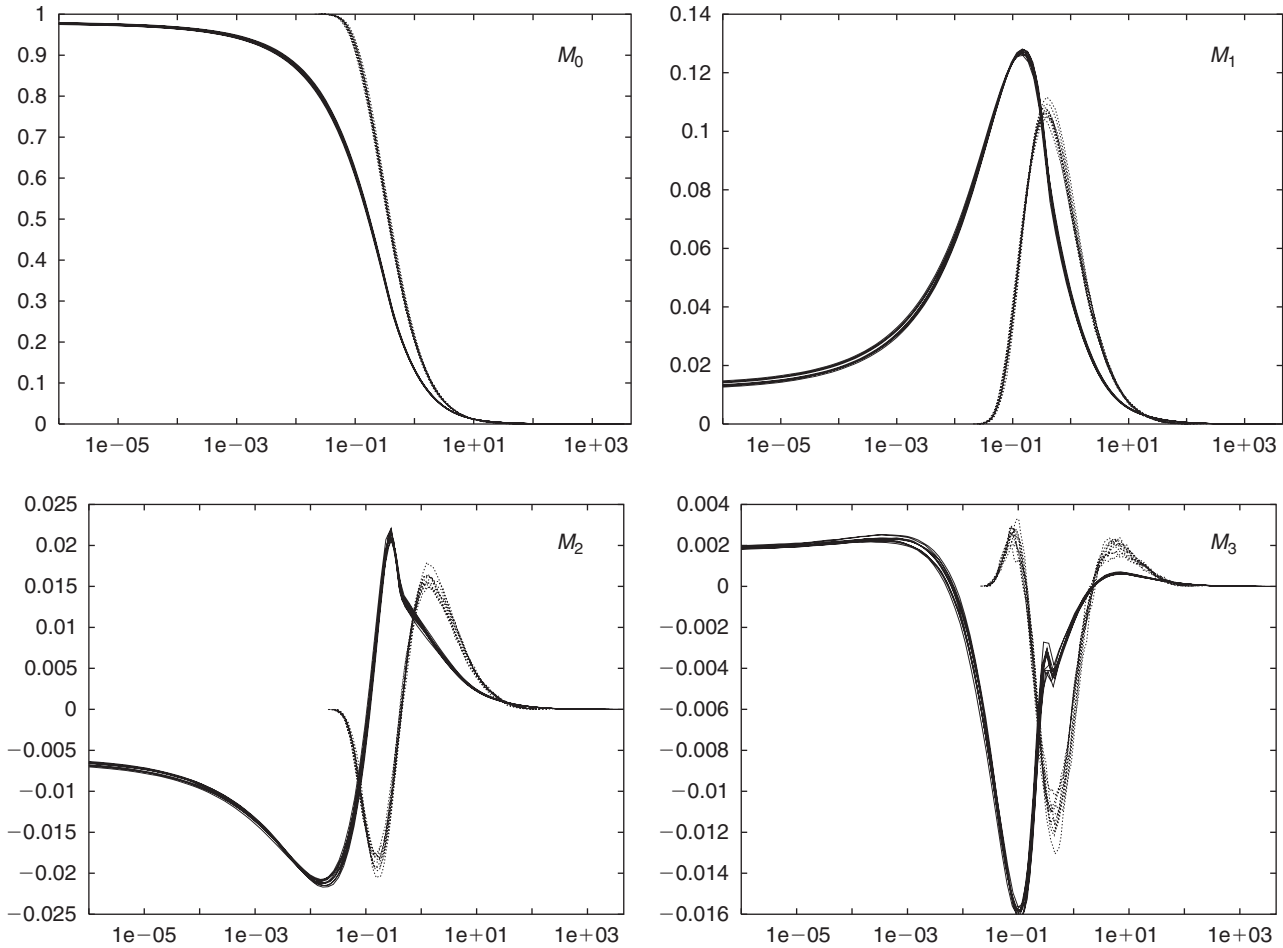
We have also computed the four scalar Minkowski functionals (MFs) of each realisation (Mecke, Buchert, & Wagner 1994). The MFs are morphological measures of the structure, sensitive to correlations of order higher than the second. They include the *genus* statistics (Melott 1990) and have been used to quantify how filamentary or sponge-like the matter distribution looks (Schmalzing et al. 1999), to study galaxy distribution in catalogues (Kerscher et al. 1997), and to address Gaussianity in the cosmic microwave background (Schmalzing & Gorski 1998).

As in deriving the power spectrum, we constructed a density field on a regular grid using the triangular shape cloud method with two different resolutions: a  $128^3$ -cell grid and a  $32^3$ -cell grid. A density threshold was introduced and the boundary surface was determined between regions with a density below the threshold and regions with a density above it. Finally, the MFs of the boundary surface were determined. The four MFs are defined as follows:

- $M_0$  = volume enclosed by the surface,
- $M_1$  = area of the surface,
- $M_2$  = integral over the surface of its mean curvature,
- $M_3$  = integral over the surface of its Gaussian curvature, which coincides with the Euler characteristic (genus):

$$M_3 = \text{number of disconnected objects} \\ + \text{number of holes} - \text{number of tunnels.}$$

Figure 8 shows the initial MFs as a function of the density threshold. The general shape of the plots can be explained qualitatively: for low thresholds, there is no bounding surface at all (due to the periodic boundary conditions), and hence  $M_0$  = total simulation volume, while  $M_1 = M_2 = M_3 = 0$ . As the threshold increases, there appear first disconnected blobs of low density regions ( $M_3 > 0$ , increasing  $M_1 > 0$ , and the boundary surface is predominantly concave:  $M_2 < 0$ ); later the blobs fuse



**Figure 9** MFs of the ten realisations of the final matter distribution ( $z = 0$ ) as a function of the logarithm of the density threshold  $\log(1 + \delta)$ , at two different spatial resolutions:  $128^3$ -cell grid (—) and  $32^3$ -cell grid (···).

together and tunnels arise ( $M_3 < 0$ ), and finally the situation reverses and one ends up with independent blobs of high density regions ( $M_3 > 0$ , decreasing  $M_1 > 0$ , and the boundary surface is predominantly convex:  $M_2 > 0$ ), until the threshold becomes larger than the maximum density ( $M_0 = M_1 = M_2 = M_3 = 0$ ). There is evidence that the zeros of  $M_3(\delta)$  are strongly correlated with the percolation thresholds of the regions above or below the density threshold (Mecke & Wagner 1991).

At the initial time, density fluctuations are small and Gaussian, which explains the symmetry of the MFs with respect to  $\delta = 0$ . However, a slight asymmetry can be detected for  $M_2$  and  $M_3$  on the  $128^3$ -cell grid: the spatial resolution is large enough that the MFs are sensitive to the finite-mass effects induced by the underlying point-particle distribution. Another difference between the two grids is the dispersion among realisations, which is larger when the spatial resolution is small. The scatter in the positions of the zeros and the values of the extrema tends to increase from  $M_0$  toward  $M_3$ ; in the worst case, the value of the maxima of  $M_3(\delta)$  has a  $1\sigma$  error  $\approx 10\%$ .

Figure 9 shows the MFs for the final matter distribution. The minimum density that can be resolved is  $1 + \delta_{\min} \approx (\text{mean interparticle distance}/\text{grid constant})^3$ ,

which is 1 for the  $128^3$ -cell grid and  $1/64$  for the  $32^3$ -cell grid. Thus, the curves below these densities are in principle not reliable.<sup>5</sup> Apart from the asymmetry around  $\delta = 0$ , one observes in general that the scatter in the ordinate direction barely changes: the value of the maximum of  $M_3(\delta)$  has an error  $\approx 15$ . The abscissa dispersion, however, is larger than at the initial time: so, for example, the zeros of  $M_3(\delta)$  in the  $32^3$ -grid have an error  $\approx 6\%$ , while the uncertainty at the initial time is just  $\approx 0.3\%$ .

#### 4 Analysis II: Dark Matter Halos

The remaining analysis is going to focus on gravitationally bound halos, identified using the bound density maxima method (BDM, Klypin & Holtzman 1997). We investigate the scatter in (large-scale) clustering patterns as well as internal properties of halos introduced by the random nature of the initial conditions.

##### 4.1 Identifying Halos

We restricted our analysis to halos with more than 100 particles ( $M_{\text{vir},\min} > 10^{12} h^{-1} M_\odot$ ). This lower mass limit

<sup>5</sup>For instance, the feature at  $1 + \delta \approx 0.4$  of  $M_2$  and  $M_3$  in the  $128^3$ -grid is likely a finite-mass effect due to isolated particles (density peaks at a density  $\approx 1$ ) in the voids of the structure.



can be used to derive a lower limit for the virial radius  $R_{\text{vir},\text{min}}$  via

$$M_{\text{vir}} = \frac{4\pi}{3} \Delta_{\text{vir}} \rho_b R_{\text{vir}}^3, \quad (15)$$

where  $\rho_b$  is the background density and  $\Delta_{\text{vir}} = 340$  for the  $\Lambda$ CDM model under consideration.

The BDM code identifies local overdensity peaks by smoothing the density field on a particular scale. The particle distribution was used to iteratively find potential halo centres as the centres of mass of 20 000 spheres with radius  $R_{\text{sphere}} \approx R_{\text{vir},\text{min}} \approx 162 h^{-1} \text{ kpc}$  centred about randomly chosen particles. Once the iteration converged for all spheres we repeated the procedure using successively smaller sphere radii down to  $70 h^{-1} \text{ kpc}$ , about three times the force resolution. For each of these halo centres we stepped out in radial bins until the density dropped below  $\rho_{\text{bin}} < \Delta_{\text{vir}} \rho_b$ . This defined the outer radius  $R_{\text{vir}}$  of the halo.<sup>6</sup> We discarded all halos with less than 100 particles within  $R_{\text{vir}}$  for the further analysis.<sup>7</sup>

#### 4.2 Mass Function of Halos

The first quantity to investigate is the mass spectrum. We calculated the cumulative mass function  $n(>M)$  for our BDM halos and compared it to the analytical prediction of Press & Schechter (1974, hereafter PS),

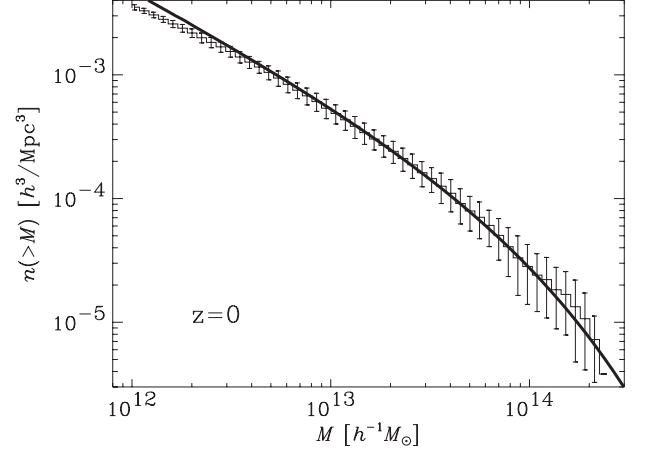
$$\frac{dn}{dM} dM = \sqrt{\frac{2}{\pi}} \frac{\langle \rho \rangle}{M} \frac{\delta_c}{\sigma_M} \left| \frac{d \ln \sigma_M}{d \ln M} \right| \exp \left( -\frac{\delta_c^2}{2\sigma_M^2} \right) \frac{dM}{M}, \quad (16)$$

where the variance  $\sigma_M$  is given by equation (6) and  $\delta_c = 1.68$ .

Figure 10 shows that the average mass function of all ten runs is in good agreement with the PS prediction, which has been noted already by several other authors (Efstathiou et al. 1988; White, Efstathiou, & Frenk 1993; Gross et al. 1998; Governato et al. 1999; Jenkins et al. 2001). This again is another indicator that the initial conditions as well as the evolution by  $N$ -body simulation are in fair agreement with theoretical predictions based on the analytical power spectrum and its time evolution. The discrepancy of the numerical  $n(>M)$  with the PS prediction at the low and high mass end of the mass function is also a well known fact (e.g. Governato et al. 1999) and not related to unreliable ICs or wrong  $N$ -body modelling. Anyway, we are more interested in the scatter stemming from the random nature of the initial conditions. We are driven by the question of to what extent a single cosmological simulation can be representative for

<sup>6</sup>If we want to identify halos-within-halos this method needs to be adjusted to account for the fact that the actual density of a satellite galaxy might not drop below  $\Delta_{\text{vir}} \rho_b$ .

<sup>7</sup>To crosscheck the completeness of our BDM halo catalogues we also performed a FOF analysis which shows a nearly 100% agreement and only an incompleteness in the BDM catalogues for halos less massive than 100 particles.



**Figure 10** Cumulative mass functions of BDM halos compared to the Press–Schechter prediction (Press & Schechter 1974). The mass functions are the average taken over all ten runs and the error bars are  $1\sigma$  errors.

the volume under investigation. We observe that the scatter gradually increases from around 4% at the very low mass end resolved to about 50% for the most massive objects found in the simulation. According to the PS prediction, the scatter due to cosmic variance should enter via the amplitude  $A_k$  predominantly, not the phases  $\theta_k$  of the ICs, equation (4). The observed increase of scatter with mass is then naturally explained also by the PS formula, given that larger masses are more sensitive to larger scales.

#### 4.3 Halo–Halo Correlation Function

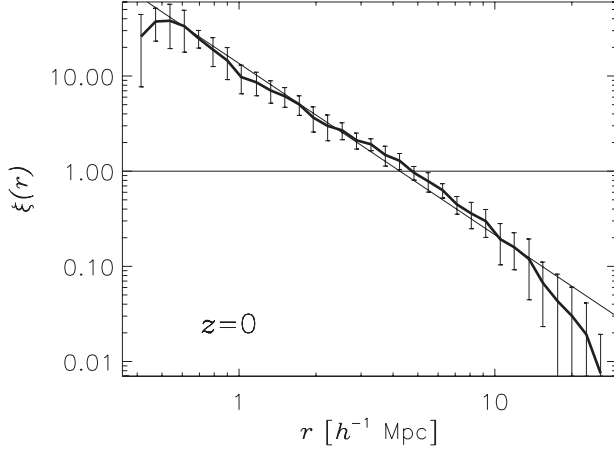
The calculation of the halo–halo two-point correlation function is based on the estimator equation (11) again. However, this time we applied it only to the 500 most massive objects in the runs, which means fixing the number density of halos to  $n_{\text{halo}} = 2 \times 10^{-3} (h^{-1} \text{ Mpc})^{-3}$ . This choice for  $n_{\text{halo}}$  restricts the masses of the objects used from  $M \sim 3 \times 10^{14} h^{-1} M_{\odot}$  down to  $M \sim 2 \times 10^{12} h^{-1} M_{\odot}$ . The result for the average taken over the ten BDM catalogues at redshift  $z=0$  is shown in Figure 11. The mean correlation function  $\xi_{\text{est}}(r)$  was fitted to a power law,

$$\xi(r) = (r_0/r)^\gamma, \quad (17)$$

over the range  $r \in [0.5, 20] h^{-1} \text{ Mpc}$  with the parameters  $r_0 = 4.26 \pm 0.44 h^{-1} \text{ Mpc}$  and  $\gamma = 1.80 \pm 0.17$ . The  $1\sigma$  errors are of the order of 10% and indicate again only a mild dependence of the halo–halo correlation function on the variance introduced by the random nature of the initial conditions. Even though the scatter for the fundamental mode is  $\approx 20\%$ , it does only marginally affect the statistical clustering properties of dark matter halos in the respective mass range.

#### 4.4 Internal Properties of the Most Massive Halo

Even though we are only resolving approximately 2% of the virial radius of the most massive particle group



**Figure 11** Two-point correlation function  $\xi(r)$  for BDM halos at redshift  $z=0$ . Error bars are again  $1\sigma$ . The thin solid line is the fit to a power law as given by equation (17).

**Table 1.** Internal (averaged) properties of the most massive halo when averaged over ten runs; the errors are the  $1\sigma$  value again

Property	Variance (%)
$M = (3.07 \pm 1.60) \times 10^{14} h^{-1} M_{\odot}$	52
$v_{\text{circ}} = (1131 \pm 199) \text{ km s}^{-1}$	18
$\sigma_v = (1172 \pm 195) \text{ km s}^{-1}$	17
$r_{\text{vir}} = (1344 \pm 163) h^{-1} \text{ kpc}$	12
$c = 4.10 \pm 0.91$	22
$\lambda = 0.033 \pm 0.018$	53
$T = 0.762 \pm 0.102$	13

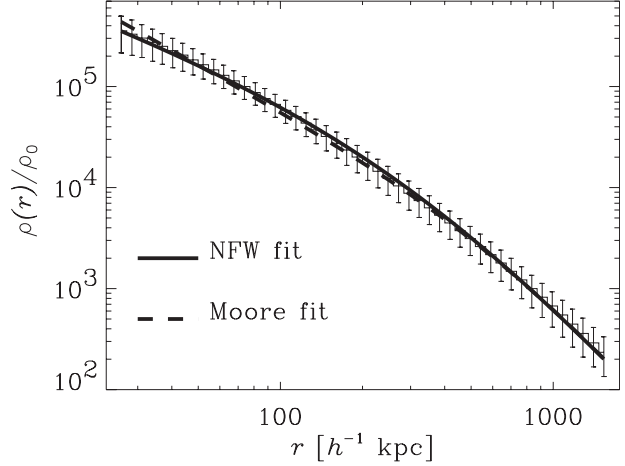
(cf. Table 1), we fitted a Navarro, Frenk, & White (NFW) profile (Navarro, Frenk, & White 1997),

$$\rho_{\text{NFW}}(r) \propto \frac{1}{r/r_s(1+r/r_s)^2}, \quad (18)$$

as well as a Moore profile (Moore et al. 1999),

$$\rho_{\text{Moore}}(r) \propto \frac{1}{(r/r_s)^{1.5}(1+(r/r_s)^{1.5})}, \quad (19)$$

to the most massive halo found in the BDM catalogues. The question we are interested in is whether the scatter due to the random nature of the initial conditions can be made responsible for the difference in the central slope of the density profile described by those two fitting formulae. And from Figure 12 we deduce that at least down to the resolved scale of 2% of the virial radius both analytical descriptions for the density profile do give indistinguishable good fits to the actual data; they both lie within the  $1\sigma$  error bars. However, we must stress that both profiles start to deviate even more strongly from each other for even smaller scales not covered by the current study. Moreover, the reduced  $\chi^2$  value for the NFW fit is marginally better than for the Moore fit, as one might anticipate from the behaviour at small  $r$  values in Figure 12.



**Figure 12** Average density profile for the most massive BDM halo with  $1\sigma$  error bars along with a NFW and a Moore profile fit to the data.

We conclude the analysis with Table 1 summarising some internal properties calculated for the most massive halo, i.e. mass  $M$ , circular velocity  $v_{\text{circ}}$ , velocity dispersion  $\sigma_v$ , virial radius  $r_{\text{vir}}$ , concentration parameter

$$c = r_{\text{vir}}/r_s, \quad (20)$$

where  $r_s$  is the scale radius derived from the fit to the NFW profile equation (18), the spin parameter

$$\lambda = J\sqrt{|E|}/(GM^{5/2}), \quad (21)$$

and the triaxiality parameter

$$T = \frac{a^2 - b^2}{a^2 - c^2}, \quad (22)$$

where  $a > b > c$  are the eigenvalues of the inertia tensor.

This table shows that the  $1\sigma$  variance for nearly all quantities is of the order of 20%, like the variance of the fundamental mode  $k_{\text{min}} = 2\pi/B$  (cf. Figure 1). Only the mass and the spin parameter show a larger scatter.

## 5 Summary and Conclusions

We present the study of ten random realisations of a density field characterised by a cosmological power spectrum  $P(k)$  at redshift  $z=50$ . These initial conditions for  $N$ -body simulations were tested with respect to their correlation properties. Recent claims by Baertschiger & Sylos Labini (2002) throw doubts on the ability of the commonly used method for generating the initial density field using particles (i.e. glass or grid preinitial distribution + the Zeldovich approximation, Eftstahiou et al. 1985) to clearly reproduce the analytical input correlations. The power spectrum  $P(k)$  and the mass variance  $\sigma_M(r)$  do not deviate from the expected behaviour (including expected departures from the desired  $\Lambda$ CDM behaviour due to finite mass and finite size effects). The estimated 2-point correlation  $\xi(r)$  is too noisy to be used as a reliable credibility check; one cannot claim either that it reproduces the desired

$\Lambda$ CDM behaviour or that it exhibits systematic deviations thereof.

These initial conditions were then evolved forward in time until redshift  $z=0$  using the publicly available adaptive mesh refinement code MLAPM (Knebe et al. 2001). This allowed us to explore the cosmic variance stemming from the random nature of the initial conditions, i.e., the scatter between different realisations of statistically identical initial conditions. We addressed the morphological properties of the matter distribution with the four Minkowski functionals as functions of a density threshold. The scatter grows in time, the one exhibiting a larger dispersion being the genus, of the order of 10% at  $z=0$ . We also investigated the internal properties of dark matter halos, which have already been shown by other groups to be profoundly influenced by the surrounding large-scale structure, which in turn is sensitive to  $k$ -modes  $\approx$  fundamental mode (Colberg et al. 1999). We find that the scatter in the properties of the most massive object(s) forming in the box is  $\sim 20\%$ , and as high as  $\sim 50\%$  for some properties such as the mass or the spin parameter.

An interesting question is whether this scatter is induced mainly by the cosmic variance of the amplitude at scales around the fundamental mode, or by the cosmic variance of the random phases. There is certainly a propagation of the error in the initial large-scale amplitude by power transfer towards smaller scales. In fact, when comparing our data to the (non-)linear fit of Peacock & Dodds (1996) for the power spectrum and to the prediction by Press & Schechter (1974) for the mass function, we find good agreement. The derivation of both results is based on the hypothesis of small influence from coupling of modes at some  $k$  to modes with larger  $k$ ; our results support this assumption, as far as the statistical estimators we probed are concerned. It would now be interesting to investigate in detail the actual influence of the large waves on the small scale structure. This would also shed some light on the credibility of running small simulation boxes to very low redshifts as already done by several groups (e.g. Dave et al. 2001; Avila-Reese et al. 2001; Gnedin 2000; Colin, Avila-Reese, & Valenzuela 2000), but we leave this to a future study.

### Acknowledgments

We would like to thank Brad Gibson for a careful reading of the manuscript and valuable comments. AK greatly acknowledges the hospitality of Rosa Domínguez Tenreiro

and Gustavo Yepes at Universidad Autónoma de Madrid where this work was started. AD thanks K. Mecke for the code to compute the Minkowski functionals.

The simulations presented in this paper were carried out on the Beowulf cluster at the Centre for Astrophysics and Supercomputing, Swinburne University. AK acknowledges the support of the Swinburne University Research Development Grants Scheme.

### References

- Avila-Reese, V., Colin, P., Valenzuela, O., D’Onghia, E., & Firmani, C. 2001, *ApJ*, 559, 516
- Baertschiger, T., & Sylos Labini, F. 2002, *EL*, 57, 322
- Colberg, J.M., White, S.D.M., Jenkins, A., & Pearce, F. 1999, *MNRAS*, 308, 593
- Colin, P., Avila-Reese, V., & Valenzuela, O. 2000, *ApJ*, 542, 622
- Dave, R., Spergel, D.N., Steinhardt, P.J., & Wandelt, B.D. 2001, *ApJ*, 547, 574
- Efstathiou, G., Frenk, C.S., & White, S.D.M. 1985, *ApJS*, 57, 241
- Efstathiou, G., Frenk, C.S., White, S.D.M., & Davis, M. 1988, *MNRAS*, 235, 715
- Gabrielli, A., Joyce, M., & Sylos Labini, F. 2002, *PhRvD*, 65, 083523
- Gnedin, N.Y. 2000, *ApJ*, 542, 535
- Governato, F., Babul, A., Quinn, T., Tozzi, P., Baugh, C.M., Katz, N., & Lake, G. 1999, *MNRAS*, 307, 949
- Gross, M.A.K., Somerville, R.S., Primack, J.R., Holtzman, J., & Klypin, A.A. 1998, *MNRAS*, 301, 81
- Hamana, T., Yoshida, N., & Suto, Y. 2001, *ApJ*, 568, 455
- Jenkins, A., Frenk, C.S., White, S.D.M., Colberg, J.M., Cole, S., Evrard, A.E., Couchman, H.M.P., & Yoshida, N. 2001, *MNRAS*, 321, 372
- Jenkins, A., et al. 1998, *ApJ*, 499, 20
- Kerscher, M., et al. 1997, *MNRAS*, 284, 73
- Klypin, A.A., & Holtzman, J. 1997, *astro-ph/9712217*
- Knebe, A., Green, A., & Binney, J. 2001, *MNRAS*, 325, 845
- Mecke, K.R., & Wagner, H. 1991, *JSP*, 64, 843
- Mecke, K.R., Buchert, T., & Wagner, H. 1994, *A&A*, 288, 697
- Melott, A.L. 1990, *PhR*, 193, 1
- Moore, B., Quinn, T., Governato, F., Stadel, J., & Lake, G. 1999, *MNRAS*, 310, 1147
- Navarro, J., Frenk, C.S., & White, S.D.M. 1997, *ApJ*, 490, 493
- Peacock, J.A., & Dodds, S.J. 1996, *MNRAS*, 280, L19 (PD96)
- Pen, U.L. 1997, *ApJL*, 490, 127
- Press, W.H., & Schechter, P. 1974, *ApJ*, 187, 425
- Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. 1992, *Numerical Recipes* (Cambridge: Cambridge University Press)
- Schmalzing, J., & Gorski, K.M. 1998, *MNRAS*, 297, 355
- Schmalzing, J., Buchert, T., Melott, A.L., Sahni, V., Sathyaprakash, B.S., & Shandarin, S.F. 1999, *ApJ*, 526, 568
- Seljak, U., & Zaldarriaga, M. 1996, *ApJ*, 469, 437
- White, S.D.M. 1996, in *Cosmology and Large-Scale Structure*, eds. R. Schaeffer, J. Silk, M. Spiro, & J. Zinn-Justin (Amsterdam: Elsevier Science), 349
- White, S.D.M., Efstathiou, G., & Frenk, C.S. 1993, *MNRAS*, 262, 1023