Setting the Stage: Structures from Gaussian Random Fields

Till Sawala,1⋆, Adrian Jenkins2, Stuart McAlpine1, Jens Jasche3, Guilhem Lavaux4, Peter H. Johansson1, Carlos S. Frenk2
1Department of Physics, Gustaf Hällströmin katu 2, University of Helsinki, Finland
2Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, United Kingdom
3The Oskar Klein Centre, Department of Physics, Stockholm University, Albanova University Center, 106 91 Stockholm, Sweden
4CNRS & Sorbonne Université, UMR7095, Institut d’Astrophysique de Paris, 75014 Paris, France

Accepted XXX. Received YYY; in original form ZZZ.

ABSTRACT
We study structure formation in a set of cosmological simulations to uncover the scales in the initial density field that gave rise to the formation of present-day structures. Our simulations share a common primordial power spectrum (here ΛCDM), but the introduction of hierarchical variations of the phase information allows us to systematically study the scales that determine the formation of structure at later times. We consider the variance in z = 0 statistics such as the matter power spectrum and halo mass function. We also define a criterion for the existence of individual haloes across simulations, and determine what scales in the initial density field contain sufficient information for the non-linear formation of unique haloes. We study how the characteristics of individual haloes such as the mass and concentration, as well as the position and velocity, are affected by variations on different scales, and give scaling relations for haloes of different mass. Finally, we use the example of a cluster-mass halo to show how our hierarchical parametrisation of the initial density field can be used to create variants of particular objects. With properties such as mass, concentration, kinematics and substructure of haloes set on distinct and well-determined scales, and its unique ability to introduce variations localised in real space, our method is a powerful tool to study structure formation in cosmological simulations.

Key words: methods: numerical – galaxies: formation – cosmology: theory, dark matter, large-scale structure of the Universe

1 INTRODUCTION
In the standard cosmological model galaxies and dark matter haloes originate from random, adiabatic density fluctuations in the Big Bang, magnified by inflation, and amplified under the force of gravity in competition with cosmic expansion. If, as inflation predicts (e.g. Linde 2005), and observations indicate (e.g. Bouchet et al. 1993; Nusser et al. 1995; Planck Collaboration 2016, 2019), the primordial density field is Gaussian, and statistically homogeneous and isotropic, its late-time power spectrum and the distribution of haloes that form in a sufficiently large volume are fully determined by the laws of physics, and the universal cosmological parameters (e.g. Peebles 1980; Bardeen et al. 1986).

Due to the random nature of the initial density field, a comparison between theoretical predictions and observations is usually done on a population level, rather than for individual objects. Constraints on the halo mass function or the matter power spectrum require large surveys, and in the case of simulations, similarly large volumes to sufficiently sample the underlying distributions (e.g. Springel et al. 2005; Klypin et al. 2011; Angulo et al. 2012). While scales much smaller than the survey or simulation size are sampled many times, scales represented by the largest haloes, or the largest modes of the density field, are sampled much more sparsely. This problem of “cosmic variance” (see, e.g. Colombi et al. 2000) is not just a question of size, however. Any particular object, observed with enough detail, is unlikely to have a closely matching counterpart in a finite simulation volume, and the search for the initial conditions that give rise to the formation of particular observed structures remains an ongoing challenge (e.g. Hoffman & Ribak 1991; Bistolas & Hoffman 1998; Yepes et al. 2014; Doumler et al. 2013; Jasche & Wandelt 2013; Hoffman et al. 2015; Jasche et al. 2015; Lavaux & Jasche 2016; Carlesi et al. 2016).

While the statistics of the halo population are fully determined by the initial power spectrum, the formation of particular objects and their characteristics also depend on the particular phase information. In a simple, monolithic collapse model (e.g. Peebles 1980), the formation of a structure would be governed only by modes with wavelengths at or above the scale of the Lagrangian region from which the structure originated. N-body simulations (e.g. Davis et al. 1985) have shown, however, that structure formation in ΛCDM is
Throughout this paper, we use physical units of mass and density, and evolved only under the effect of gravity. In this section, we describe the creation of our initial conditions, the set-up of the simulations, and the identification of structures. Additional information about the parametrisation of the primordial density field, and the Panphasia white noise field that is used in its construction, can be found in Jenkins (2013) and Jenkins & Booth (2013).

2 METHODS

The results in this paper are based on cosmological “dark matter only” N-body simulations, i.e. both baryons and dark matter are subsumed into a single type of simulation particle, and evolved

1 Throughout this paper, we use physical units of mass and density, and comoving physical units of length, unless otherwise specified.
2.1.2 Contribution to the Variance from Individual Levels

To illustrate the relative importance of individual layers of octree functions to the \( \Lambda \)CDM overdensity field, we first consider the fractional contribution of each level of octree functions to the total variance of the overdensity field when smoothed with a spherical top-hat function of radius \( R \), containing mass \( M \), at the mean density of the universe. This ratio is given by:

\[
\frac{\sigma_L^2}{\sigma^2} = \frac{\int d^3k \ W^2(kR) \left( \frac{j^2_{\text{lin}}(k\Delta_L)}{j^2_{\text{lin}}(2k\Delta_L)} P_{\text{lin}}(k) \right)}{\int d^3k \ W^2(kR) P_{\text{lin}}(k)},
\]

where \( W(kR) \equiv (\sin(kR) - kR \cos(kR))/(kR)^3 \) is the spherical top-hat window function in \( k \)-space, \( P_{\text{lin}}(k) \) is the \( \Lambda \)CDM linear power spectrum, \( \Delta_L \) is the cell size of the octree at level \( L \), and the functions \( j_{\text{lin}} \) are given by:

\[
\begin{align*}
\frac{j_{\text{lin}}(k\Delta_L)}{j_{\text{lin}}(2k\Delta_L)} &= \frac{(2l + 1)(2m + 1)(2n + 1)j_l(k\Delta_L)j_m(k_x\Delta_L)j_n(k_z\Delta_L)}{2^2}.
\end{align*}
\]

The three functions \( j_{l,m,n} \), on the right-hand side, are Spherical Bessel functions of the first kind, and \( k_x, k_y, k_z \) are the Cartesian components of the wave vector \( k \).

The phase information for level \( L \) is defined as that given by the octree functions that fully occupy octree cells at level \( L - 1 \), and therefore contribute phase information that can be represented as sets of disjoint polynomials filling the octree cells at level \( L \).

In Fig. 1 we plot the ratio \( \sigma_L^2/\sigma^2 \) against the spherical top-hat filter mass, \( M \), and size of a spherical perturbation, \( \lambda \), related via \( M = \bar{\rho} \pi R^3/3 \), where \( \bar{\rho} = \Omega_m \times 3H_0^2/(8\pi G) \), and \( r = \lambda/2 = \pi/k \), for a series of single octree levels, each shown by a solid line. For each level there is a range of top-hat filter masses and wavelengths where a particular level contributes most to the total variance. We call these scales \( M_{\text{max}L} \) and \( \lambda_{\text{max}L} \), respectively.

For comparison, the dashed curves show the fractional contribution to the variance if, instead of setting octree functions to zero except in a single octree layer, we do the equivalent transformation assuming we use Fourier modes to represent the phase information. More precisely, we set all Fourier modes to zero outside the factor of two range of \( 1/\sqrt{2} < k\Delta_L < \sqrt{2} \).

The peaks of the dashed and solid curves line up quite well, and the shape of the sets of curves are similar. This indicates that for quantities such as the variance of the overdensity field smoothed with a top-hat, we can establish a close correspondence between octree layers and sharp-\( k \) space shells in Fourier space. We expect to get very similar results to those we present in this paper, had we chosen instead only to work with Fourier modes.

2.1.3 The Cut-Off Scale

If we set to zero all the octree functions occupying cells at level \( L \) and above (so that the white noise field is approximated by disjoint polynomial functions occupying the octree cells at level \( L \)), the resulting, truncated power spectrum has a high-\( k \) cut-off of the form:

\[
\langle p_{\text{lin}}^2(k) \rangle = P_{\text{lin}}(k) \sum_{l,m,n=0,1} j_{\text{lin}}^2(k\Delta_L).
\]

The high-\( k \) cut-off is due to the summation term in brackets on the right-hand side of Eqn. (3) which tends to one in the limit of small \( k \), and to zero for large \( k \).

In Fig. 2 we illustrate how the full \( \Lambda \)CDM dimensionless power spectrum is built up from successive higher ‘layers’ of octree functions by plotting Eqn. 3 for a \( \Lambda \)CDM power spectrum for several values of \( L \). We can see that the \( \Lambda \)CDM linear power spectrum is approximated better and better as we add successive layers of octree function phase information. Truncating the octree representation for all cells higher than a given level produces a relatively sharp

---

2. http://virgo.dur.ac.uk/
Figure 2. The reproduction of the ΛCDM dimensionless power spectrum (blue solid line) as a function of wavenumber, \( k \), bottom axis, and wavelength, \( \lambda \), top axis, with increasing accuracy. Individual black lines show the truncated power spectra given by Eqn. 3, successively including the phase information for octree levels up to the indicated level \( L \). The dashed grey line indicates a reduction in \( \Delta^2 \) by a factor of 4. Faint vertical lines mark its intercept with the truncated power spectra, allowing us to define a cut-off wavenumber, \( k_{\text{cut}} \), and cut-off scales, \( \lambda_{\text{cut}}(L) \) and \( M_{\text{cut}}(L) \) for each octree level.

cut-off in the spherically averaged power spectrum. We can define an associated cut-off wavenumber, \( k_{\text{cut}}(L) \), by determining where \( P_{\text{lin}}(k) \) falls by a factor of two in amplitude, or four in power, below the ΛCDM linear power spectrum.

### 2.1.4 Length Scales and Physical Quantities

In Table 1 we give a set of conversion factors which relate the level, \( L \), to physical quantities. The first column gives the octree level. Level 23 is the deepest level and hence the smallest scale at which we sample the Panphasia field for this project. The second column lists the side length of the octree cells themselves. As the 100 cMpc side length of the simulation volume is represented by 12 cells at level 12, the length-scale at level \( L \) is \((100/3)^2 14^L \) cMpc.

| \( L \) | \( \Delta_{\text{cell}} \) | \( \lambda_{\text{max},2} \) | \( \log_{10} \left( \frac{M_{\text{max},2}}{M_\odot} \right) \) | \( \lambda_{\text{cut}} \) | \( \log_{10} \left( \frac{M_{\text{cut}}}{M_\odot} \right) \) |
|---|---|---|---|---|---|
| 23 | 0.065 | 0.12 | 5.6 | 0.051 | 7.3 |
| 22 | 0.13 | 0.27 | 6.1 | 0.101 | 8.2 |
| 33 | 0.26 | 0.61 | 7.5 | 0.202 | 9.1 |
| 20 | 0.52 | 0.12 | 8.4 | 0.405 | 10.0 |
| 19 | 1.04 | 0.27 | 9.5 | 0.81 | 10.9 |
| 18 | 2.08 | 0.61 | 10.5 | 1.62 | 11.8 |
| 17 | 4.17 | 1.22 | 11.4 | 3.24 | 12.7 |
| 16 | 8.33 | 2.73 | 12.5 | 6.48 | 13.6 |
| 15 | 16.67 | 6.11 | 13.5 | 12.97 | 14.5 |
| 14 | 33.33 | 13.68 | 14.6 | 25.94 | 15.4 |
| 13 | 66.67 | 30.63 | 15.6 | 51.88 | 16.3 |
| 12 | 133.33 | 61.11 | 16.5 | 103.78 | 17.2 |

L: octree level, \( \Delta_{\text{cell}} \): size of the octree cell at level \( L \), \( \lambda_{\text{max},2} \): characteristic wavelength, where the fractional contribution to the variance from level \( L \) is maximal, \( M_{\text{max},2} \): mass of a spherical top hat of diameter \( \lambda_{\text{max},2} \), \( \lambda_{\text{cut}} \): wavelength at which power is reduced by four when power at levels \( L \) and above are set to zero, \( M_{\text{cut}} \): mass of a spherical top hat of diameter \( \lambda_{\text{cut}} \).

L = 20 contribute most at \( \lambda_{\text{max},2} = 0.12 \) cMpc, the power already falls to 1/4 at \( \lambda_{\text{cut}} = 0.41 \) cMpc if the power spectrum is truncated at \( L = 20 \). This is due to the fact that the information from \( L = 20 \) already contributes significantly to the full ΛCDM power spectrum at this scale, in addition to smaller contributions from \( L > 20 \). While \( \lambda_{\text{max},2} \) thus locates the peak contribution from a single level, for the effect of variations at and above a certain scale on structure formation, \( \lambda_{\text{cut}} \) proves the most useful quantity. In Section 4.2, we also define a new length scale based on the existence of unique haloes across variations at a given level.

### 2.1.5 Panphasia

In addition to defining the \( S_8 \) octree basis functions, Jenkins (2013) also defined a single extremely large ‘public’ realisation of a Gaussian white noise field, called ‘Panphasia’. The Panphasia field is an octree with 50 levels, more than enough to encompass all the phase information of all existing cosmological simulations. By design the phase information in Panphasia can be computed rapidly at any location of the field and at any depth in the octree. We will take our phase information for this paper from the Panphasia field.

All the simulations for this paper are of a \( 100^3 \) cMpc\(^3 \) volume. We define a reference set of phase information which is the phase information used for the Eagle Project flagship \( 100^3 \) cMpc\(^3 \) volume (Schaye et al. 2015). The phase information for this simulation occupies a very small region of the entire Panphasia field. The reference simulation phase information comes from a cubic patch of dimension \( 12^3 \) at the twelfth level octree (so at this level the whole Panphasia field consists of \( 2^{12} = 2048 \) cells on a side). We will use the symbol, \( L \), usually with a subscript to denote the level of phase information in the Panphasia field. The zero point for the octree levels is arbitrary and simply follows as a consequence of a choice for the reference phases made by Schaye et al. (2015).
2.1.6 Variations as Shifts in Panphasia

As illustrated in Fig. 3, we can conveniently introduce random variations to the initial density field as coordinate shifts in Panphasia. The left panel shows three adjacent sub-volumes of Panphasia, in blue, yellow, and red, respectively. Each sub-volume contains completely independent phase information, i.e. completely independent regions of the white noise field, for the same simulation volume. The next three panels, to the right, show several possible variants. Assuming that the blue Panphasia region contains the full phase information of the Reference simulation from levels $L_{\text{min}}$ to $L_{\text{max}}$, variants 1 and 2 differ from the Reference simulation, and from each other, at levels $L_{\text{min}+2}$ and above. Variant 3 shares levels $L_{\text{min}}$ to $L_{\text{min}+2}$ with the Reference simulation, levels $L_{\text{min}+3}$ with the second variant, and differs from all other simulations at level $L_{\text{min}+4}$.

Provided that there is significant large-scale power, initial conditions with different low-level phase information result in the formation of different objects, independent of shared high-level phase information. The Reference simulation, and the first and second variants, share the same amount of phase information ($L_{\text{min}}$ to $L_{\text{min}+2}$), so it is expected that, statistically, the structures formed in each will be equally similar to one another. Since the levels of Panphasia are completely independent, the choice of “Reference” among the three is arbitrary. If two simulations share the same large-scale phase information, the statistical similarity in the structures formed depends on the smallest scale down to which the phase information is shared. While Variants 2 and 3 also share phase information at level $L_{\text{min}+4}$, the fact that they differ at $L_{\text{min}+3}$ means that they have no additional similarity. Out of all illustrated volumes, the Reference simulation and Variant 3 share phase information down to the smallest scale, so it is expected that they will have the greatest similarity in the structures formed.

We label a set of simulations that differ from the phase information of the Reference simulation from level $L$ to $L_{\text{max}}$ as variants VL, and identify individual volumes that employ a shift by $i$ from level $L$ to $L_{\text{max}}$ as $V_{L,i}$. For example, $V_{L,16}$ is the set of variants that differ from the Reference simulation at level $L = 16$ and above, and $V_{185}$ is the individual variant that differs from the Reference simulation at level $L = 18$ and above by a shift by 5. Variants with multiple shifts, e.g. by $i$ from level $L$ to level $M - i$ and by $j$ from level $M$ to level $L_{\text{max}}$, where $L < M \leq L_{\text{max}}$, are labelled $V_{L,i}/M_{j}$, etc. Hence, $V_{185/215}$ is the individual variant that employs a shift by 2 at levels 18 to 20, and a shift by 5 at levels 21 and above.

2.2 Simulations and structure finding

All simulations presented here assume a ΛCDM cosmology, with parameters $h = 0.6777$, $n_s = 0.9611$, $\sigma_8 = 0.8288$, $\Omega_0 = 0.307$, $\Omega_b = 0.0483$, and $\Omega_{\Lambda} = 0.693$. They are set up in a volume of $100^3$ cMpc$^3$, using $N = 384^3$ ($5.3 \times 10^8$) particles, giving a particle mass of $7.4 \times 10^8 M_{\odot}$, a mean interparticle separation of 260 ckpc, and a comoving softening length of 13 ckpc throughout.

We use the IC_Gen initial conditions code and the methods described in the papers Jenkins (2010, 2013) to make second-order Lagrangian perturbation theory initial conditions for a starting redshift of 127. We used a 1536$^3$ Fourier mesh to generated all of the initial conditions. The Nyquist frequency of this Fourier mesh is a factor of four smaller than the particle Nyquist frequency. We set all Fourier modes to zero if the magnitude of their wavenumber equals or exceeds the one-dimensional particle Nyquist frequency.

The simulations are run using P-GADGET-3, a TreePM code based on the publicly available code GADGET-2 (Springel 2005). In total, we have performed 469 simulations: one simulation with phase information identical to the 100 cMpc EAGLE volume of Schaye et al. (2015), hereafter called "Reference", and 39 "variant" simulations, for each of the 12 levels from 12 to 23. At every level, there are thus 40 simulations including the Reference simulation, which are equidistant in their white noise fields.

Overdensities and self-bound structures are identified using the FoF (Davis et al. 1985) and SUBFIND (Springel et al. 2001) algorithms, respectively. Throughout this paper, unless otherwise mentioned, the term "halo" refers to a self-bound structure, as identified by SUBFIND, and we limit some of our analysis to only central

---

**Figure 3.** Illustration of displacements in the Panphasia white noise field. In all panels the region of Panphasia represented in the simulation volume is indicated by thick lines. In the left panel, adjacent regions of Panphasia, displaced by 1 and 2 in one dimension are shown in red and yellow, respectively. The second and third panels (Variants 1 and 2) illustrate shifts at level $L_{\text{min}+2}$ by 1 and 2, respectively. The fourth panel, Variant 3, represents a shift by 1 at level $L_{\text{min}+3}$ and another arbitrary shift at level $L_{\text{min}+4}$. Any combination of shifts is possible, and all integer shifts result in independent white noise fields on the respective levels.
haloes, i.e. the most massive self-bound structures within their FoF groups.

3 GLOBAL RESULTS

In this section, we present the overall results of our simulations, and the effect of the varying the initial conditions on the $z = 0$ density field, the matter power spectrum, and the abundance of dark matter haloes of different mass.

3.1 Density Fields

In Fig. 4, we visually compare the structures at $z = 0$ formed in the Reference simulation (top row) to those formed in variants with independent WNs at and above levels 22, 20, 18, and 16, respectively. From left to right, the columns show the projected dark matter density in boxes of side length 100, 33 and 10 Mpc, respectively. Shades of grey indicate similar projected density in the Reference and the variant simulations, shades of blue or purple indicate higher densities in the Reference or variant simulations, respectively.

In the right column, differences from the Reference simulation can already be perceived at level 22. While nearly all identifiable haloes can be matched by eye, some low mass haloes appear slightly displaced, often by less than the size of the halo. At this level of variation, scales visible in the middle and left panels appear almost identical. For variants at level 20, most haloes in the right column are offset, but can still be matched across simulations by eye. Differences are also apparent in the middle column, where some lower mass haloes now show a noticeable displacement. At $L = 18$, differences are noticeable in all three panels. In the right column, all haloes appear visibly displaced, and many low mass haloes can no longer be matched by eye, while in the left column, displacements are still mostly below the size of the identifiable haloes. At $L = 16$ all differences are enhanced; while the right hand panel shows similar amounts of structure in both simulations, most individual objects can no longer be identified and appear at random. The middle panel still shows some correlation between the position of the more massive groups, as well as filaments, but only the largest haloes still appear in dark grey, indicating that they are displaced by less than their size.

We will examine the changes to individual, matched objects more rigorously in Section 4. As an example, we will also discuss changes to a single, cluster-mass halo in more detail in Section 6.

3.2 Power Spectrum

In Fig. 5, we show the matter power spectrum, $P(k)$, of our simulations measured at $z = 0$. We define

$$P(k) = \frac{1}{V} \langle \delta^2(k) \rangle^2$$

as the volume-averaged power spectrum, where $\delta_k$ is given by the three-dimensional Fourier transform of the density perturbation field, $\delta(x) = \rho(x)/\langle \rho \rangle - 1$, over the simulation volume.

While all simulations are set up with an identical input power spectrum, each one only contains a finite volume, and hence each mode is sampled only a finite number of time. The black solid line in Fig. 5 shows the result of the Reference simulation, and coloured lines show the results of the 39 variants at each level from 12 to 23, offset for visual clarity. The Reference result is also repeated as a dashed line with every set. In addition, thick grey lines show the average power spectra off all variants at a given level. It can be seen that, for small variations (e.g. for variations from level 23), all variants have a nearly identical power spectra on all scales, and follow all the peculiar features of the Reference simulation. As the scale of variations increases, differences between individual variants can be seen, but not at all scales: very small scales (large $k$) are sampled so well within each volume that differences between the variants are averaged out, while very large scales (small $k$) are not yet affected at moderate scales of variation. The large-scale limit of scatter in the matter power spectra grows with decreasing $L$. For larger variations, it can also be seen that the average power spectrum is much smoother than any individual power spectrum, effectively sampling a larger volume. The difference between levels 12 and 13, however, is minimal. It can be seen that individual lines for individual volumes nearly match one another. This is due to the fact that there is very little power at level 12, subject to the mean-density constraint of the simulation volume. By coincidence, the Reference simulation is quite close to this average on all scales.

3.3 Halo abundance

In Fig. 6, we compare the halo mass functions in terms of $M_{200,\text{crit}}$ of our simulations to the analytic mass function of Jenkins et al. (2001), calculated for the same cosmology. Due to the small sample size, the Reference simulation (thick black solid line), measured in a single, 100$^3$ Mpc$^3$ volume, differs slightly from the mass function of Jenkins et al. (2001) (thick black dashed line) at high masses. Similarly to Fig. 5, coloured lines show the results of the 39 variants at each level from $L = 14$ to $L = 23$. For visual clarity, each set of lines is offset, and the mass functions of Jenkins et al. (2001) and the Reference simulations are repeated as thin dashed and solid black lines with each set.

It can be seen that, for small variations, the halo mass functions very closely follow that of the Reference simulation. The scatter continuously increases as the scale of the variations increases. Only by level 15 ($\Lambda = 25.6$ Mpc) does the scatter among the mass functions for individual volumes become large enough to erase the particular features inherited from the Reference simulation. Similar to the variation of the power spectrum, beyond level 13, there is very little additional variation.

It can also be seen that the scatter among the mass functions

\begin{table}[h]
\centering
\begin{tabular}{cccccc}
\hline
Level & \multicolumn{4}{c}{\textbf{Number of Haloes}} & \textbf{Ref} & \textbf{21689} & \textbf{2602} & \textbf{343} & \textbf{21} \\
\hline
13 & 21723 ± 48 & 2613 ± 9 & 343 ± 1 & 21 ± 0 & 22 & 21683 ± 51 & 2597 ± 6 & 345 ± 2 & 21 ± 0 & 21 & 21676 ± 78 & 2613 ± 16 & 341 ± 3 & 21 ± 0 & 20 & 21704 ± 81 & 2629 ± 18 & 344 ± 6 & 21 ± 1 & 19 & 21790 ± 63 & 2698 ± 28 & 347 ± 8 & 21 ± 1 & 18 & 21833 ± 118 & 2662 ± 46 & 356 ± 12 & 23 ± 2 & 17 & 21825 ± 257 & 2703 ± 28 & 328 ± 8 & 28 ± 2 & 16 & 21711 ± 214 & 2674 ± 66 & 327 ± 22 & 27 ± 3 & 15 & 21403 ± 266 & 2655 ± 80 & 318 ± 13 & 32 ± 5 & 14 & 21705 ± 292 & 2732 ± 100 & 322 ± 16 & 29 ± 5 & 13 & 21554 ± 184 & 2677 ± 87 & 318 ± 25 & 29 ± 5 & 12 & 21523 ± 193 & 2682 ± 89 & 318 ± 25 & 30 ± 6 & 11 & 21407 ± 251 & 2662 ± 97 & 318 ± 30 & 30 ± 6 & 10 & 21511 ± 250 & 2681 ± 89 & 317 ± 25 & 30 ± 6 & 9 & 21498 ± 249 & 2655 ± 80 & 318 ± 25 & 30 ± 6 & 8 & 21486 ± 250 & 2664 ± 97 & 318 ± 30 & 30 ± 6 & 7 & 21489 ± 260 & 2671 ± 87 & 318 ± 25 & 30 ± 6 & 6 & 21457 ± 223 & 2645 ± 71 & 318 ± 25 & 30 ± 6 \\
\hline
\end{tabular}
\caption{Median and standard deviation in the number haloes of a given mass across the simulations of a given level.}
\end{table}
Figure 4. Dark matter density at $z = 0$. From left to right, columns show boxes of width and depth 100, 33 and 10 Mpc, respectively. The top row shows the simulation with the Reference phase information, while the following rows show the effect of randomising phase information at and above levels 22, 20, 18, respectively. On each panel, shades of grey indicate high projected density in both the Reference and variant simulations, blue indicates higher density in the Reference simulation, purple indicates higher density in the variant simulations.
for each set is greater at higher mass. This may appear counterintuitive, given that (as we discuss in Section 5.1), individual, higher mass objects are less strongly affected by changes to the primordial density field at a given scale. However, the scatter in the mass function has different origins at different masses: at the high mass end, where the number of haloes is low, the scatter is due to a change in the mass of individual objects, while at the low mass end, it is due to the change in the number of independent objects, but on scale with very small sampling noise.

Table 2 gives an overview of the median number of haloes of different masses, and the associated standard deviation, across the 39 simulations for each level from 12 to 23. It is worth noting that, for sufficiently large haloes, or sufficiently small scale variations, the standard deviation in halo number, \(\sigma_N\), is below even the value of \(\sqrt{N}\) expected for a random process without any variation in the bias. For example, the average number of haloes of \(10^{13} M_\odot\) is \(\sim 2700 \sim 525\), but the scatter at \(L = 20\) is only 18. This indicates that the scatter is due primarily to a change in the mass of the same haloes found across different simulations. For larger variations, the scatter typically rises above \(\sqrt{N}\), which can be attributed to different bias in each volume (White & Rees 1978; Cole & Kaiser 1989).

Fig. 7 presents the same information visually. It can be seen that, while variations in the initial density field at \(L = 19\) leads to the formation of independent \(10^{14} M_\odot\) haloes, their population is so well sampled in the 100Mpc volume that their number has less than a 1% scatter at any level. At the other end of the mass range most individual \(10^{14} M_\odot\) haloes exist across all simulations at \(L = 17\), albeit with a small change in mass. The simulation volume, however, is not large enough to sample them accurately in every volume. Averaging over all variants at levels 12 or 13, we can see that the Reference simulation contains slightly fewer than the expected \(30 \pm 5\) haloes of mass \(10^{14} M_\odot\).

### 4 INDEPENDENT HALOES

In the previous section, we have compared populations of haloes among different simulations. We now turn our attention to individual haloes. As we already discussed, if the variations in the initial conditions between two simulations are small enough, the same haloes form in both. Here, we investigate which scales in the initial density field determine the existence of unique haloes, and how the properties of individual haloes change subject to variations on smaller scales.

In order to address these questions, we need to match haloes across simulations. Following Springel et al. (2008), we use the fact...
that all of our simulations start from identical glass files, with particles whose IDs encode their initial, unperturbed Lagrangian coordinates along a Peano-Hilbert curve. Fig. 8 illustrates the matching procedure for a halo identified in a snapshot of the Reference simulation, represented by the leftmost slice. In a first step, we identify the halo’s 50 most bound particles (or all particles, if fewer than 50), and use their IDs to determine their Lagrangian origin. This is represented by the second slice from the left, where particles identified in the previous step occupy a finite volume, indicated by the grey circle.

In the next step, we examine all haloes in the corresponding snapshot in one of the variant simulations, each represented by the five coloured slices to the right. We select haloes whose masses are within a factor of three of that in the Reference simulation. If a halo contains a large fraction of the 50 particles identified in the previous step, by definition, its Lagrangian region of origin overlaps with that of the halo in the Reference simulation. In the example shown here, the first (red) slice, third (blue) and fourth (green) slice each contain a halo that fulfils the mass criterion and contains at least 1/5 of the particles of the halo in the Reference simulation. These haloes, which have grown to a similar mass from similar regions of origin, are considered matches. Conversely, the second (yellow) and fifth (purple) slice do not contain matching haloes in this example.

We note that this procedure is not completely symmetrical under exchange of the Reference and the variant simulations. However, we have tested that our results do not vary qualitatively when, at a given level, one of the variants is chosen as the Reference instead.

As expected, the matching rate is highest for high masses and high level (small-scale) changes in the initial density field, and decreases when the scale of changes increases relative to the size of the haloes. However, we find that even for low-mass haloes in simulations that share no phase information, the matching rate increases when the scale of changes increases relative to the size of high level (small-scale) changes in the initial density field, and decreases when the scale of changes increases relative to the size of high level (small-scale) changes in the initial density field.

4.1 Halo Identity Across Simulations

The possibility that two similar haloes can exist in two volumes that share no phase information leaves the tantalising question: when are two haloes genuinely identical? It appears that simply asking that they consist of the same particles is not sufficient; instead, we are looking for haloes that are formed for the same physical reasons.

While we could modify our matching criteria, we cannot discriminate ab initio between a genuine match (one where the halo pair has formed because of the common phase information) and a merely coincidental one. However, if spurious matches occur purely by chance, and the rate of those matches is less than 1/2, the probability for at least $N/2$ spurious matches to the same halo in $N$ variant simulations decreases with $N$. Conversely, if genuine matches are found with a probability above 1/2, the probability for $N/2$ genuine matches to $N$ variants simulations increases with $N$. Consequently, for sufficiently large $N$, genuine matches have a high probability to be identified in more than half of the variants, and haloes that are matched to more than half of the variants have a high likelihood of being genuine matches.

In Fig. 9, on each panel, we show the multiple-matching rate, $f_N$, as a function of mass, for 1 to 39 possible matches between the Reference simulation and all $N = 39$ variants at each level. Shades of red denote the fraction of haloes with 1 to 19 matches; shades of blue correspond to 20 to 39 matches. The thick black line shows the fraction of haloes with 20 matches. It can be seen that the fraction of haloes above $10^{10.5}M_\odot$ matched at least once is over 95%, almost independently of mass and level. However, as expected for purely chance events, at lower masses and larger scale variations, the number of multiple matches rapidly decreases, and as expected, the fraction of haloes matched at least half of the time...
match, N

Figure 9. Fraction of haloes with multiple matches, \( f_{\text{match},N} \), between the Reference simulation and the 39 variant simulations at each level from 16 to 21, as a function of halo mass in the Reference simulation. Shades of red and blue indicate 1 to 19, and 21 to 39 matches, respectively, while the black line indicates 20 matches. As expected, the matching rate is highest for high masses and high level (small scale) variations, and decreases towards lower masses and higher level variations. It is worth noting that even at scales that are completely uncorrelated, many haloes can be matched by chance. However, the number of multiple matches decreases sharply.

4.2 Existence of Unique Haloes

The matching by "majority vote" introduced in the previous paragraph allows us to define a new criterion for the existence of unique haloes; as illustrated in Fig. 10, we say that a halo exists at and above level \( \ell_E \) if it can be matched to more than half of the simulations that randomly vary the initial density fields at levels above \( \ell_E \). In other words, a halo exists at a scale \( \ell_E \); because, at this scale, the initial density field contains the necessary information for the formation of this particular halo.

In Fig. 11, we show the fraction of unique haloes that can be matched to more than half of the variants, \( f_{\text{match},1/2} \), as a function of level, \( \ell \), and of the corresponding cut-off wavelength, \( \lambda_{\text{cut}}(\ell) \). Circles show the results measured in our simulations, dashed lines of corresponding colours show two-parameter logistic fits of the form

\[
 f_{\text{match},1/2}(\lambda_{\text{cut}}, M) = \frac{1}{1 + e^{-a_M(\lambda_{\text{cut}}-\lambda_{E,M})}},
\]

where \( a_M \) and \( \lambda_{E,M} \) are free parameters fit separately at each mass, over the domain indicated by the extent of the dashed lines.

For each mass, we find a similar behaviour, with values of \( a_M \) in the range \( 8 - 10 \). Moreover, we find a regular scaling of \( \log(\lambda_{E,M}) \propto \log(M) \). Consequently, the solid lines show a global fit to Eqn. 4, with \( a_M = a = 9 \), and

\[
 \log_{10}(\lambda_{E,M}/\text{cMpc}) = 0.34 \log_{10}(M/M_\odot) - 3.85. \tag{5}
\]

Dashed lines give a closer match to the individual halo mass ranges for which they are fitted independently with two free parameters each. However, the solid lines give a close fit to the entire data set with only three free parameters, \( a \) and the two coefficients in Eqn. 5 that determine \( \lambda_{E,M} \).

The phase information of the initial density field has to be defined at least down to a scale where \( \lambda_{\text{cut}} = \lambda_{E,M} \) in order to not only create the same number of haloes of a given mass at \( z = 0 \), but to specify the formation of unique haloes. We find that this scale, \( \lambda_{E} \), is \( \sim 30\% \) smaller than \( \lambda_{\text{cut}} \) for haloes of the same mass. As examples, we find that unique haloes of \( 10^{12} \) and \( 10^{14}M_\odot \) exist at \( \lambda_{E} = 1.7 \) and 8.1 cMpc, respectively, while the cut-off scales for \( 10^{12}M_\odot \) and \( 10^{14}M_\odot \) are \( \lambda_{\text{cut}} = 2.4 \) and 11.6 cMpc, respectively.

5 THE ORIGIN OF HALO PROPERTIES

Having defined the scales that determine the existence of particular haloes, we now turn to the changes seen in the properties of individual haloes that are matched across simulations due to variations on smaller scales. In particular, we will examine the mass and concentration of individual haloes, and their position and velocities.

At this stage, it is important to remind ourselves that, while the Reference simulation plays a special role in identifying matches, it is only one of many possible realisations. At every level, it shares the same amount of phase information with any of the variants, as they share with one another. Furthermore, any halo identified in the Reference simulation is only one possible realisation of that halo. If we consider that the possible random variations on scales below \( \ell_E \) define the space of possible halo properties, we can consider the halo of the Reference simulation as one random sample of this space, not guaranteed to be at its centre. In measuring the variation...
in properties of individual haloes, we therefore do no compare the different realisations of a halo to the Reference simulation. Instead, we compute a median value from all realisations at a given level, and analyse the scatter among individual samples.

5.1 Mass

Fig. 12 shows the change in mass of individual, matched haloes, relative to the median mass across variants, and as a function of median mass, for variations at levels 18 to 23. On every panel, each individual halo can appear from 20 to 40 times; haloes with fewer than 20 matches are excluded by the "majority vote" criterion described in Section 4. The range on all panels covers mass ratios between 1 : 3 and 3 : 1. It is worth remembering that matches with mass ratios outside this range relative to the Reference simulation are excluded. While haloes can, in principle, have mass ratios of nearly 1 : 9 relative to the median mass, the paucity of points approaching the limits of the range indicates that such large deviations in mass are very rare.

We find that, at a given halo mass and level of variation, the halo masses of variants are approximately log-normally distributed. Consequently, we quantify the scatter\(^3\) in mass across the variants for each halo as

\[ \sigma_M^* = \sigma(\log_{10} M). \] (6)

The red bands on each panel of Fig. 12 indicate \( \sigma_M^* \), in bins equally spaced in \( \log(<M>) \). It can be seen that, at each level, the scatter increases with decreasing halo mass. Comparing different panels, it can also be seen that the relative variance in mass increases as the scale of the variation in the initial density field is increased from \( L = 23 \) to \( L = 18 \). For haloes of a given mass, the mass varies more for lower level variations.

This behaviour is summarised across masses and levels in Fig. 13, which shows \( \sigma_M^* \) as a function of level and \( \lambda_{\text{cut}} \). Different coloured points show the scatter measured from our simulations for haloes of different mass, from \( 10^{11}M_\odot \) (blue) to \( 10^{14}M_\odot \) (grey). To each set of data points, we have fitted linear relationships of the form

\[ \sigma_M^* = a \lambda_{\text{cut}} + b_M \] (7)

which are shown in Fig. 13 by dashed lines of corresponding colours, with coefficients \( a_M \) and \( b_M \) depending on mass.

Because haloes of a given mass drop out of existence below the level \( L_{E}(M) \), given by Eqn. 5, data points below \( L_{E}(M) \) are excluded from the fit. We also exclude haloes at \( L = 23 \), where we find a slight upturn in \( \sigma_M^*(L) \) relative to the fit at all masses. We attribute this to the finite spatial resolution of our simulations.

We also find a universal relation for the mass scatter, with a value of \( a = 0.15 \) for all mass bins, and a mass-dependence for \( b_M = 4.6 - 0.445 \log_{10}(M/M_\odot) \), i.e.

\[ \sigma_M^* = 4.6 + 1.15 \lambda_{\text{cut}} - 0.445 \log_{10}(M/M_\odot). \] (8)

Fits to this universal relation are shown by solid lines in Fig. 13.

From the observed regularity, we expect that the relations can be extrapolated both to higher mass haloes and to smaller scales. However, it should not be extrapolated to levels below the existence scale for a given halo mass.

Comparing \( \sigma_M^*(L) \) to the existence scale, \( L_{E} \), we find that the scatter in halo mass typically reaches \( \sigma_M^* \sim 0.23 \), or a factor of \( \sim 1.7 \) in mass before haloes drop out of existence. Scatters in mass of \( \sim 10\% \) (\( \sigma_M^* = 0.041 \)) or \( \sim 1\% \) (\( \sigma_M^* = 0.0043 \)) are found for variations at 2 and 5 levels above the existence level \( L_{E} \), respectively.

5.2 Concentration

In addition to the mass, a second parameter is required to characterise a ΛCDM halo whose density is described by an NFW profile (Navarro et al. 1996),

\[ \rho = \frac{\rho_s}{(r/r_s)(1 + r/r_s)^2}. \] (9)

\(^3\) We refer to the "scatter" of a quantity, \( x \), as either the standard deviation, \( \sigma_x = \sigma(x) \), or the standard deviation of its logarithm, \( \sigma^*_x = \sigma(\log_{10}(x)). \)
where $r_s$ and $\rho_s$ are the scale radius and characteristic density, respectively. The mass is commonly complemented by the concentration, $c$, as a second parameter, defined through the equation

$$\delta_c = \frac{200}{3} \frac{c^3}{\ln(1 + c) - c/(1 + c)},$$  \hspace{1cm} (10)$$

where $\delta_c$ is the characteristic overdensity.

Following Springel et al. (2008), we calculate $c$ from the measured values of the maximum circular velocity, $v_{\text{max}}$, and its corresponding radius, $r_{\text{max}}$, from which the overdensity inside $r_{\text{max}}$ can be obtained via:

$$\delta_v = \frac{\rho(r_{\text{max}})}{\rho_{\text{crit}}} = 2 \left( \frac{v_{\text{max}}}{H_0 r_{\text{max}}} \right)^2.$$

For an NFW halo, the characteristic overdensity, $\delta_c$, of Eqn. 10 is related to the overdensity inside $r_{\text{max}}$, $\delta_v$, via the relation

$$\delta_c = \frac{\rho_s}{\rho_{\text{crit}}} = 7.213 \delta_v.$$

We can thus calculate the concentration parameter, $c$, of each halo from the measured values of $r_{\text{max}}$ and $v_{\text{max}}$ by computing $\delta_v$ and $\delta_c$, and solving Eqn. 10 for $c$.

Because stripping of satellites affects $r_{\text{max}}$ disproportionately relative to $v_{\text{max}}$, Eqn. 10 is not expected to hold for satellites, and we limit our analysis to centrals only. Furthermore, measuring the concentration of a halo requires sufficient numerical resolution to resolve the density inside $r_{\text{max}}$. As Power et al. (2003) showed, even in an accurate simulation, this is sensitive to a halo’s particle number, and we limit our analysis of concentration to haloes containing at least 1000 particles, or $7.4 \times 10^{11} M_\odot$.

Neto et al. (2007) showed that concentration parameters for relaxed haloes of a given mass are approximately log-normal distributed, with $c_\text{M}^* \sim 0.11$ at $10^{12} M_\odot$. Fig. 14 shows the relative scatter in concentration of matched central haloes, as a function of median mass, and for variations at levels 18 to 23. The red band shows $c_\text{M}^*$.

In contrast to the scatter in halo mass, there is a noticeable scatter in concentration already at the smallest scale of variation. For example, the scatter in concentration of individual $10^{12} M_\odot$ haloes at $L = 22$ is 8%, while the scatter in mass is < 1% for the same variation. This reflects the fact that the scale that determines the concentration of a halo, $r_{\text{max}}$, is much smaller than its total size.

Figure 12. Mass ratio between the mass of individual, matched central haloes in all 39 variant simulations at a given level, and their average mass across all 40 simulations, as a function of average mass. The red band shows the 1-sigma scatter. Vertical dashed lines on panels at $L = 18$ and 19 indicate the minimum mass of haloes that exist at these level.

Figure 13. Median scatter in log halo mass, $\sigma^*_M$, for matched haloes of different mass (indicated by different colours), as a function of level of randomisation in the initial density field ($L$, upper x-axis), and corresponding wavelength $\lambda_{\text{cMpc}}$. Only bold data points are included in the fits. Dashed lines show independent, power-law fits for each mass bin, solid lines show a universal power-law fit for all masses.
and we find a much greater scatter in $r_z$ than in $r_{200}$ for the same haloes.

While the measurable scatter in total mass is bounded only by the matching criteria, the scatter in concentration is naturally bounded from above by the narrow range of concentrations for $\Lambda CDM$ haloes. As expected, we find that for large variations, the scatter of matched haloes is similar to the scatter of independent haloes reported by Neto et al. (2007).

Fig. 15 shows $\sigma_r^c$ for different masses, and different scales of variation. As expected, we find that the scatter increases with decreasing halo mass, and with the scale of variations. As shown by dashed lines. Over the range we can resolve, and up to the maximum value of $\sim 0.1$, $\sigma_r^c$ appears to scale linearly with $\lambda_{\text{cut}}$, and also show a regular scaling with mass. However, our simulations lack the dynamic range necessary to extrapolate universal scaling relations.

### 5.3 Position

Fig. 16 shows the relative displacement, $\Delta r$, of individual, matched haloes at $z = 0$ from their median position, as a function of their median mass, and for variations at levels 18 to 23. Unlike in Fig. 12, the range on each panel is adjusted to include 99% of points in every case.

We calculate the scatter in position as

$$\sigma_r = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}$$

(11)

where $r(x, y, z)$ are the Cartesian coordinates across the variants of the halo, and $\sigma_x$, $\sigma_y$, and $\sigma_z$ are the 1D-dispersions in position.

In Fig. 16, $\sigma_r$ is shown by a red line. At each level, the average displacement increases with decreasing halo mass: for fixed variations of the initial density field, lower mass haloes are displaced more than high mass haloes. Comparing different panels, it can also be seen that the displacement of haloes of the same mass increases as the scale of the variation in the initial density field is increased: haloes of a given mass are more displaced for larger scale variations. At $L = 19$ and $L = 18$, the mass-dependence appears to flatten at low masses. However, in both cases, this coincides with the limit in halo mass for unique haloes to exist, as indicated by the vertical dashed lines; haloes below this limit have a low probability to form across variants.
Figure 16. Displacement of matched central haloes in the variant simulation relative to the median position for each halo, as a function of median halo mass, for different levels (indicated on each panel). The red lines show one standard deviation in position among variants. Vertical dashed lines on panels at $L = 18$ and $19$ indicate the minimum mass of haloes that exist at these level. On each panel, the range extends to the 99th percentile of all haloes shown.

Figure 17. Scatter of halo positions, $\sigma_r$, for central haloes of different mass (indicated by different colours), and as a function of level of randomisation in the initial density field ($L$, upper x-axis), and corresponding cut-off wavelength ($\lambda_{\text{cut}}$, lower x-axis). Only bold data points are included in the fits. Dashed lines show independent, power-law fits for each mass bin, solid lines show a universal power-law fit for all masses.

The dependence of average displacement on halo mass and level is illustrated in Fig. 17. Different coloured symbols show the scatter measured from our simulations for haloes of different mass, from $10^{11} M_\odot$ (blue) to $10^{14} M_\odot$ (grey). We find the average displacement of haloes in each mass bin to be a power-law function of $\lambda_{\text{cut}}$ of the form

$$\log_{10}(\sigma_r/\text{cMpc}) = \alpha_M \log_{10}(\lambda_{\text{cut}}/\text{cMpc}) + b_M.$$  \hspace{1cm} (12)

Dashed lines Fig. 17 show separate fits at each mass, when the same limits to the domain as in Section 5.1 are applied. As in Eqs. 4 and 7, we find a self-similar behaviour, with a universal slope of $a = \alpha_M = 1.5$, and a mass-dependence of $b_M = 3.55 - 0.31 \log_{10}(M/\text{M}_\odot)$. Solid lines show fits with these parameters of $a$ and $b_M$ at every level.

The increase in the average displacement of haloes with the scale of the perturbations follows from the fact that the position of an object depends on the gravitational potential of the surrounding density field, and greater variation in surrounding structure leads to greater variation in the potential, and thus the position of the halo. The mass-dependence at a given scale of variation follows from the fact that less massive objects are sensitive to smaller external perturbations of the potential than more massive ones. Comparing the mass- and scale-dependencies of $\sigma_r$ to those of $\sigma_v^*$, we find a stronger scale-dependence for $\sigma_r$, and a stronger mass-dependence for $\sigma_v^*$.

5.4 Velocity

By analogy to Fig. 16, Fig. 18 shows the scatter in velocity of individual, matched haloes at $z = 0$, as a function of their median mass, for variations at levels 18 to 23. The range on each panel is adjusted to include 99% of points in every case.

The velocity dispersion, $\sigma_v$, of individual haloes is defined as the dispersion in peculiar velocities of its matched variants,

$$\sigma_v = \sqrt{\sigma(v_x)^2 + \sigma(v_y)^2 + \sigma(v_z)^2},$$  \hspace{1cm} (13)

where $\sigma(v_x)$, $\sigma(v_y)$ and $\sigma(v_z)$ are the corresponding 1D-velocity dispersions.

The median velocity scatter among matches, as a function of halo mass and level, is illustrated in Fig. 19. Different coloured symbols show the scatter measured from our simulations for haloes...
Setting the Stage: Structures from Gaussian Random Fields

6 VARIATIONS OF A SINGLE HALO

In this section, we show how the introduction of random variations on different scales can create targeted variations of a particular object, either very similar to, or very different from the original object. In addition, we show how introducing an additional set of higher level variations can create small additional perturbations to an existing variant.

As an example, we choose the most massive halo of the Reference simulation at $z = 0$, with a mass of $2.2 \times 10^{14} M_{\odot}$, whose density map is shown in the top left panel of Fig. 20, and which is comparable in mass to the Virgo cluster (Urban et al. 2011).

Galaxy clusters are a frequent target for zoom-in simulations, e.g. by Eke et al. (1998), and more recently by Borgani et al. (2002); Kay et al. (2004); Nagai et al. (2007); Martizzi et al. (2016); Bahé et al. (2017). Schaller et al. (2015) and Barnes et al. (2017) both analysed clusters from the same Eagle volume that defines our ref-
An inherent challenge in these studies is that the largest clusters are, by definition, rare objects in any cosmological volume, limiting the predictive power of a simulation. For example, Schaller et al. (2015) conclude that the presence of cores in observed clusters is an outstanding problem that requires a larger samples of simulation counterparts, which our method of generating variants of existing objects may be able to provide.

In Fig. 20, we show the density profile of the same cluster. As expected, from the top centre and top right panel, it can be seen that compared to the Reference simulation, V18, its central density and mass of the main halo change throughout. At levels 21 and above, the density in the centre of the main halo remains similar, but the positions of the larger substructures change, resulting in some scatter in the outer density profile. Below level 20, the central density and mass of the main halo change throughout.

At levels L = 20, 18, and 16, the scatter in mass is 1%, 10%, and 60%, respectively, while the scatter in concentration is 9%, 20%, and 35%. Variations below level 20 yield clusters with very similar mass and concentrations, but different individual subhaloes; variations at level 18 yield clusters with very similar mass, but different concentration (and different subhaloes, apart from the largest ones), and variations at level 16 result in clusters with different total masses and concentrations, and completely different members.

6.1 Simple Variations

In Fig. 21, we show the density profile of the same cluster in the Reference simulation (black line) and, offset for clarity, for up to 39 matches at each level from 16 to 23. At small scales (V23 and V22), there is very little change in the density profile, both for the main halo, as well as for the identifiable substructures. At levels 21 and 20, the density in the centre of the main halo remains similar, but the positions of the larger substructures change, resulting in some scatter in the outer density profile. Below level 20, the central density and mass of the main halo change throughout.

At levels L = 20, 18, and 16, the scatter in mass is 1%, 10%, and 60%, respectively, while the scatter in concentration is 9%, 20%, and 35%. Variations below level 20 yield clusters with very similar mass and concentrations, but different individual subhaloes; variations at level 18 yield clusters with very similar mass, but different concentration (and different subhaloes, apart from the largest ones), and variations at level 16 result in clusters with different total masses and concentrations, and completely different members.

6.2 Higher Level Perturbations

As discussed in Section 2.1.1 and illustrated in the rightmost panel of Fig. 3, we can also combine multiple shifts at different levels in the same simulations. As an example, we have made a set of additional simulations, in which we shift the phase information by 1 at levels 18 to 20, and by 1-19 at levels 21 and above. We label this new set of simulations ‘V18\(_{1}/21_{1-19}\)’ and note that V18\(_{1}/21_{1}\) is identical to V18\(_{1}\).

In Fig. 20, we show the changes resulting from these higher level perturbations, by comparing the projected density in higher level perturbations at L = 21 to both the reference simulation, and to the variant V18\(_{1}\) of the same cluster. As expected, from the top centre and top right panel, it can be seen that compared to the Reference simulation, V18\(_{1}\) and V18\(_{1}/21_{2}\) show the same amount
of difference, while V18, shown on the bottom left, is much more similar to the Reference simulation.

The bottom middle and right panel show the density relative to V18 for variants V18/21 and V18/21, respectively. While their differences relative to the reference simulation are similar in magnitude to that of V18, their differences relative to V18 are much smaller, and similar in magnitude to that between V21 and the reference simulation.

Similar to Fig. 21, in Fig. 22, we show the density profiles of the cluster in the 19 variants V18 - 19 (red lines), and the 19 variants V18/21 - 19 (blue lines), with offsets for clarity. It can be seen that the scatter among the density profiles for V18/21 - 19 is much reduced compared to the scatter among V18 - 19, and furthermore, that the individual variants of V18/21 - 19 scatter around the variation V18 (dashed black line), rather than around the Reference simulation (solid black line).

Variations of a single object by random variations of the initial density field at scales below the existence scale can be used deliberately, to facilitate the study of rare objects. Variations just below the existence scale could be used to create a diverse set of massive clusters using only zoom-in simulations, without requiring to first simulate a much larger simulation volume. By introducing secondary variations at level 21 to an existing variant at level 18, we can create multiple, smaller scale perturbations around existing objects. This offers many possible applications: for example, from a cluster with moderate concentration, we can first create a variant with high concentration, and then create multiple high-concentration clusters with separate satellite populations.

It is also worth noting that we have applied the shift across the entire volume. This is not necessary, the real-space localisation of our basis allows us to independently vary spatial sub-volumes of the random density field. In a forthcoming paper, we will show how this technique can be used to create a faithful reproduction of the Local Group embedded within the observed constraints. In particular, it allows the generation of accurate Local Group candidates through a series of localised, small-scale perturbations, without affecting the nearby large-scale structure constraints.

7 SUMMARY AND DISCUSSION

Using cosmological N-body simulations, we have identified the scales of the initial density field that statistically determine the formation of haloes of different mass and their abundance. We have also determined the scales responsible for the existence of particular haloes, and explored how the properties of individual haloes change, as a result of variations on smaller scales.

We have defined a criterion for the existence of unique haloes across simulation volumes: a halo is said to exist at a certain scale, L, if the information in the primordial density field is sufficient for more than half of the variations at smaller scales still to result in the formation of the same halo. In Eqn. 5, we have parametrised the existence scale as a function of halo mass.

Beyond the mere existence of haloes, we have also quantified changes in their properties such as mass, concentration, position and velocity, and find power power-law relationships between the scale of the initial variation and the scatter in each of these properties. Furthermore, over the range 10^{11} - 10^{14} M_{\odot} in halo mass, we find that for each property, a single relationship can be used, when scaled by halo mass, reflecting the self-similarity of hierarchical growth from ΛCDM initial conditions.

The change in mass shows the strongest mass-dependence. By contrast, there is only a very weak mass dependence to the change in velocity. We attribute this to the fact that, while the mass accretion is set on the same scale that determines the total halo mass, the velocities of haloes are set by the tidal field on larger scales. The small remaining mass dependence can be attributed to the degree to which the halo itself influences its own tidal environment.

Taking a cluster-mass halo as an example, we also show that, by varying different scales in the initial conditions, variants of the same object can be produced. Small-scale variations will result in the same cluster, but substructures of different mass, different orbits, or in different phases of their orbit. Larger scale variations will change the overall mass and concentration of the cluster itself. In zoom simulations of particular objects, this can be used to estimate the expected, random change in properties, arising from the fact that the initial density field is more finely sampled. Moreover, changes to the white noise field can also be introduced deliberately to explore regions of the parameter space, which can be useful in the study of rare objects.

Our method bares similarity to the "genetically modified haloes" approach of Roth et al. (2016), which extends the recursive Hoffman-Ribak algorithm (Hoffman & Ribak 1991) to compute the changes to the initial density field required for the formation of a structure with desired properties, along with the necessary corrections to preserve the nature of the random field. Extending the biological metaphor, our method relies on evolution through fully random mutations, rather than geneticmodification. However, because it allows independent control of the scale of variations and of their spatial extent, it can be used to explore the vast space of possibilities very efficiently, while preserving the Gaussian nature of the initial random field at all times.

Our method can be naturally extended in several ways: multiple levels of the white noise field can be varied independently; different spatial regions of the white noise field can be varied independently, and variations smaller in amplitude than a full substitution can be used, all while preserving the Gaussian nature of the initial conditions. In a forthcoming paper, we will use this method to produce a faithful representation of the Local Group embedded in the observed large scale constraints of the Local Universe.

ACKNOWLEDGEMENTS

TS is an Academy of Finland Research fellow. TS and SM are supported by the Academy of Finland grant 314238. PHJ acknowledges the support by the European Research Council via ERC Consolidator Grant (no. 818930). CSF acknowledges support from European Research Council (ERC) Advanced Investigator grant DMDAS (GA 786910). This work was also supported by the Consolidated Grant for Astronomy at Durham (ST/L00075X/1).

This work made use the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure. Simulations were also performed on facilities hosted by the CSC - IT Centre for Science, Finland.
REFERENCES

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Aragon-Calvo M. A., 2016, MNRAS, 455, 438
Bahé Y. M., et al., 2017, MNRAS, 470, 4186
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Barnes D. J., et al., 2017, MNRAS, 471, 1088
Bertschinger E., 2001, ApJS, 137, 1
Bistolas V., Hoffman Y. 1998, The Astrophysical Journal, 492, 439A-451
Borgani S., Governato F., Wadsley J., Menci N., Tozzi P., Quinn T., Stadel J., Lake G., 2002, MNRAS, 336, 409
Bouchet F. R., Strauss M. A., Davis M., Fisher K. B., Yahil A., Huchra J. P., 1993, ApJ, 417, 36
Carlesi E., et al., 2016, Monthly Notices of the Royal Astronomical Society, 458, 900A-911
Colombi S., Zampa I., Jenkins A., Colberg J., 2000, MNRAS, 313, 711
Davis M., Efstathiou A., Peiris H. V., White S. D. M., 1985, ApJ, 292, 371
Dounler J., Gottlöber S., Hoffman Y., Courtou H., 2013, Monthly Notices of the Royal Astronomical Society, 430, 912A-923
Efstathiou G., Davis M., White S. D. M., Frenk C. S., 1985, ApJS, 57, 241
Eke V. R., Navarro J. F., Frenk C. S., 1998, ApJ, 503, 569
Grand R. J. J., et al., 2017, MNRAS, 467, 179
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hoffman Y., Ribak E., 1991, ApJ, 380, L5
Hoffman Y., Courtou H., Tully R. B., 2015, MNRAS, 449, 4494
Jasche J., Wandelt B. D., 2013, MNRAS, 432, 594
Jasche J., Leclercq F., Wandelt B. D., 2015, J. Cosmology Astropart. Phys., 2015, 036
Jenkins A., 2010, MNRAS, 403, 1859
Jenkins A., Booth S., 2013, Panphasia: a user guide (arXiv:1306.5771)
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
Katz N., Quinn T., Bertschinger E., Gelb J. M., 1994, MNRAS, 270, L71
Kay S. T., Thomas P. A., Jenkins A., Pearce F. R., 2004, MNRAS, 355, 1091
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Kay S. T., Thomas P. A., Jenkins A., Pearce F. R., 2004, MNRAS, 355, 1091
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Lavaux G., Jasche J., 2016, MNRAS, 455, 3169
Linde A., 2005, arXiv e-prints, pp hep-th/0503203
Martizzi D., Hahn O., Wu H.-Y., Evrard A. E., Teyssier R., Wechsler R. H., 2016, MNRAS, 459, 4408
Nagai D., Vikhlinin A., Kravtsov A. V., 2007, ApJ, 655, 98
Navarro J. F., Eke V. R., Frenk C. S., 1996, MNRAS, 283, L72
Neto A. F., et al., 2007, MNRAS, 381, 1450
Nusser A., Dekel A., Yahil A., 1995, ApJ, 449, 439
O’Leary R. M., Quinn M., 2012, ApJ, 760, 4
Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press
Planck Collaboration 2016a, A&A, 594, A17
Planck Collaboration 2019, arXiv e-prints, arXiv:1905.05697
Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14
Roth N., Pontzen A., Peiris H. V., 2016, MNRAS, 455, 974
Salmon J., 1996, ApJ, 460, 59
Sawala T., et al., 2016, MNRAS, 457, 1931
Schaller M., et al., 2015, MNRAS, 452, 343
Schaye J., et al., 2015, MNRAS, 446, 521
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V., et al., 2005, Nature, 435, 629
Springel V., et al., 2008, MNRAS, 391, 1685
Urban O., Werner N., Simionescu A., Allen S. W., Böhringer H., 2011, MNRAS, 414, 2101
Wang J., Bose S., Frenk C. S., Gao L., Jenkins A., Springel V., White S. D. M., 2019, arXiv e-prints, arXiv:1911.09720
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341
Yepes G., Gottlöber S., Hoffman Y., 2014, New Astronomy Reviews, 58, 1

APPENDIX A: PANPHASIA DESCRIPTORS USED

The Reference simulation in this paper uses the same phase information as the 100 cMpc Eagle simulation (Schaye et al. 2015), whose Panphasia phase descriptor is:

[Panph1,L16, (31250, 23438, 39063), S12, CH1050187043, EAGLE_L0100_VOL1]

The reference phase information is defined using a $12 \times 12 \times 12$ root cell at level $L = 16$. At level $L$ for example the reference phases in the $x$-direction are taken from the range of twelve cells $31250 - 31261$, and similarly $23438-23443$ and $39063-39074$ in $y$ and $z$ respectively.

To create variants of the phase information we extract the phase information from neighbouring regions of the Panphasia field. Because a white noise field is uncorrelated it is sufficient to use regions that are shifted by multiples of 12 cells at level 16, in the $x$, $y$ or $z$ directions.

From this, the phase information for volumes $V_{Lj}$ is constructed by applying an integer spatial shift of $(\Delta x, \Delta y, \Delta z)$, in units of the box size, equivalent to 12 cells in the positive $x$-direction at level 16 of the octree, or 100 Mpc, and similarly for $y$ and $z$.

$$\Delta x = (j \mod 10),$$
$$\Delta y = ((j - \Delta x) \mod 100)/10,,$$
$$\Delta z = ((j - 100\Delta y - \Delta x) \mod 1000)/100.$$

For example, $V_{2124}$ is constructed using shifts of $\Delta x = 4$ and $\Delta y = 2$.

For every variant, we shift up to 10 levels at $L_{min}$ and above by multiples of the box size. The cell size at each level is the ratio between the box size and the root cell, $l_{cell} = 100/12^{L-16}$ cMpc. In order to shift the phase information by the box size, at each level, a shift by $\Delta N_{i} = \Delta x \times 12 \times 2^{3-L}$ is required. For example, $V_{2124}$ corresponds to the following cell shifts in Panphasia:

$$\Delta N_{x} = 4 \times 12 \times 2^{21-16} = 1536$$
$$\Delta N_{y} = 2 \times 12 \times 2^{21-16} = 768$$
$$\Delta N_{z} = 0$$

$$\Delta N_{x} = 4 \times 12 \times 2^{22-16} = 3072$$
$$\Delta N_{y} = 2 \times 12 \times 2^{22-16} = 1536$$
$$\Delta N_{z} = 0$$

$$\Delta N_{x} = 4 \times 12 \times 2^{33-16} = 6144$$
$$\Delta N_{y} = 2 \times 12 \times 2^{33-16} = 1536$$
$$\Delta N_{z} = 0$$

This paper has been typeset from a TeX/LaTeX file prepared by the author.