Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **501,** 4759–4776 (2021)
Advance Access publication 2021 January 16

doi:10.1093/mnras/staa3568

Setting the stage: structures from Gaussian random fields

Till Sawala,[1]★ Adrian Jenkins,[2] Stuart McAlpine,[1] Jens Jasche,[3] Guilhem Lavaux,[4]
Peter H. Johansson[1] and Carlos S. Frenk[2]

[1]*Department of Physics, University of Helsinki, Gustaf Hällströmin katu 2, 00560 Helsinki, Finland*
[2]*Institute for Computational Cosmology, Durham University, South Road, Durham DH1 3LE, UK*
[3]*The Oskar Klein Centre, Department of Physics, Stockholm University, Albanova University Center, SE-106 91 Stockholm, Sweden*
[4]*Institut d'Astrophysique de Paris, CNRS and Sorbonne Université, UMR7095, F-75014 Paris, France*

Accepted 2020 October 13. Received 2020 September 23; in original form 2020 March 10

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 Λ cold dark matter, 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 parametrization 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 localized in real space, our method is a powerful tool to study structure formation in cosmological simulations.

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

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, Dekel & Yahil 1995; Planck Collaboration XVII 2016; Planck Collaboration IX 2020), 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).

Because of 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, Trujillo-Gomez & Primack 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; Doumler et al. 2013; Jasche & Wandelt 2013; Yepes, Gottlöber & Hoffman 2014; Hoffman, Courtois & Tully 2015; Jasche, Leclercq & Wandelt 2015; Carlesi et al. 2016; Lavaux & Jasche 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 Λ cold dark matter (ΛCDM) is 'bottom up', and large structures form in part through the merger of smaller ones, causing even smaller scale modes to affect the formation and the properties of more massive haloes.

The fact that distinct constituents of the halo population originate from independent scales in the initial density field have been pointed out, for example, by Aragon-Calvo (2016), who used simulations with shared large-scale modes in order to create an ensemble of

---

★ E-mail: till.sawala@helsinki.fi

© 2021 The Author(s)
Published by Oxford University Press on behalf of Royal Astronomical Society

simulations. Notably, and analogous to our Sections 3.3 and 3.2, they show how a set of semi-independent realizations with some shared phase information can overcome the statistical limitations created by single sample.

In this paper, we systematically investigate which scales in the initial density field are responsible for the formation of haloes at later times. Starting from the density field of the 100^3 cMpc^3 EAGLE simulation volume[1] (Schaye et al. 2015), we generate a sequence of simulations that systematically introduce random variations of the white noise field on increasingly larger scales. At each scale, we compare the resulting simulations in terms of the evolved density field, the population of haloes, the existence of individual haloes, and the variation in halo properties.

In particular, we uncover the scales in the initial density field that contain the information responsible for the non-linear formation of individual haloes. We also examine how the properties of haloes of different masses change when the density field is perturbed on scales below this *existence* scale. Finally, we demonstrate how our method of introducing hierarchical, random perturbations to existing density fields can be used deliberately to create variations of simulations and simulated objects, including with variations in mass, concentration, and kinematics. This makes it a particularly powerful tool for future zoom and constrained simulations, allowing to efficiently explore the parameter space of possible initial conditions that give rise to the formation of haloes with particular observed properties. We will explore the full potential of this method in a forthcoming paper.

This paper is organized as follows. In Section 2.1, we describe our method for parametrizing the phase information in an octree basis, and the PANPHASIA white noise field. In Section 2.2, we outline the set-up of the simulations used in this paper, our way of identifying haloes, and of matching objects across simulations. Section 3 presents the global results of our simulations, in terms of the density field and its power spectrum, in Section 3.1, and the halo population, in Section 3.3. In Section 4, we discuss the formation, and variation in properties, of individual objects. We present a definition of the identity of particular haloes across simulations in Section 4.1, which allows us, in Section 4.2, to study the scales in the initial density field that determine the existence of particular haloes. In Section 5, we study the variation of halo properties: mass (Section 5.1), concentration (Section 5.2), position (Section 5.3), and velocity (Section 5.4). We present variations of a particular cluster-mass halo in Section 6, and conclude with a summary and an outlook to future work in Section 7.

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 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 parametrization 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).

2.1 Initial conditions

A natural way to describe the primordial Gaussian density or displacement field in a cosmological simulation of a cubic region with periodic boundary conditions is a Fourier representation, introduced for cosmological simulations by Efstathiou et al. (1985), and employed by many subsequent initial condition generators (e.g. Katz et al. 1994; Bertschinger 2001; Springel et al. 2005; Jenkins 2010; Hahn & Abel 2011; O'Leary & McQuinn 2012).

In the standard ΛCDM model, it is assumed that the initial density fluctuations after inflation are Gaussian. The statistical properties of the overdensity field, $\delta(\boldsymbol{x})$, and its Fourier transform, $\delta(\boldsymbol{k})$, for a volume, V, are then completely defined by the one-dimensional linear power spectrum, $P_{\mathrm{lin}}(k)$:

$$P_{\mathrm{lin}}(k) = \frac{1}{V}\langle|\delta_k|^2\rangle. \qquad (1)$$

To create a set of ΛCDM initial conditions, it is necessary to specify the cosmological parameters, the dimensions of the periodic region, the power spectrum, and the phase information, which can be encapsulated as a realization of a Gaussian white noise field (Salmon 1996). Gaussian white noise fields are particularly convenient to work with numerically: for example, their two-point autocorrelation functions are zero for any non-zero lag. Consequently, the values for an unconstrained white noise field can be set by any high-quality pseudo-random number generator. As Salmon (1996) also pointed out, not only does this offer a simple way to produce multiscale random fields, it is also straightforward to include linear constraints: the corresponding ΛCDM linear overdensity field for a given linear matter power spectrum is given by the convolution, in real space, of the white noise 'phase information' field with a specific spherically symmetric window function, computed from a one-dimensional integral over the linear matter power spectrum.

2.1.1 Octree basis

Jenkins (2013) introduced a way of constructing multiscale real-space white noise fields based on an octree decomposition of a cubic period volume. In this formalism, the Gaussian white noise field is built hierarchically from a linear superposition of octree basis functions.

The individual basis functions are zero outside of the particular cubic cell they occupy, and orthogonal to each other, even when completely or partially overlapping. Consequently, an unconstrained Gaussian white noise field can be created by choosing the octree function amplitudes from a pseudo-random sequence of uncorrelated Gaussian variables of zero mean and unit variance.

The white noise field can be refined at any location by adding information from higher (or deeper) levels of the octree. The white noise field is then specified in each cell at the highest level as an appropriate polynomial of the local Cartesian cell coordinates. The octree functions themselves are not polynomials, but an octree function occupying any parent cell is represented by eight distinct piecewise polynomials, with each one filling one of the parent's eight child cells.

It proves convenient in this paper to use these octree functions to represent the phase information in all the simulations we present. The primary rationale for this choice, however, will be made evident in subsequent papers. Our ultimate goal is to construct multiscale initial conditions that also satisfy linear constraints, derived for example from reconstructions of the large-scale structure around the Milky Way. Here, the octree decomposition is crucial, as it allows for the introduction of localized changes to the density field, independently from the external constraints.

---

[1] Throughout this paper, we use physical units of mass and comoving physical units of length (denoted as cMPc) and density, unless otherwise specified. At $z = 0$, we use Mpc or kpc as a shorthand.

There are an infinite number of possible sets of orthogonal octree basis function, using different polynomial functions to represent the field within the cells of the octree. The degree to which the density information is attached to a particular level of the octree depends on the choice of basis. Throughout this paper we will use the S_8 octree basis functions developed by Jenkins (2013). The S_8 function set have been used in making initial conditions for many cosmological simulations of the Virgo Consortium,[2] from modelling galaxy clusters (Bahé et al. 2017) via galaxy groups (Sawala et al. 2016) and individual galaxies (Grand et al. 2017), down to simulating the smallest cold dark matter (CDM) dark haloes at the present day (Wang et al. 2020).

We use the S_8 octree functions for practical reasons as they are well established and have been shown to be suitable for zoom simulations in the sense that that the large-scale properties of resimulated haloes closely match the originals for a reasonable computational cost. This is in contrast to the simpler S_1 octree basis functions that show poor stability in the properties of haloes with increasing resolution at the same computational cost (Jenkins 2013). We have not explored other choices for this paper. To mitigate against our particular choice of basis functions when presenting our results, we introduce later in this subsection, length scales and mass scales that take into account of our choice of basis function in addition to the physical sizes of the octree cells.

2.1.2 Contribution to the variance from individual levels

To illustrate the relative importance of individual layers of octree functions to the ΛCDM 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 and mean density, containing mass M at the mean density of the universe. This ratio is given by

$$\frac{\sigma_L^2}{\sigma^2} = \frac{\int \mathrm{d}^3\boldsymbol{k}\, W^2(kR)\left(j^2_{lmn}(\boldsymbol{k}\Delta_L) - j^2_{lmn}(2\boldsymbol{k}\Delta_L)\right) P_{\mathrm{lin}}(k)}{\int \mathrm{d}^3\boldsymbol{k}\, W^2(kR) P_{\mathrm{lin}}(k)},$$

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

$$j_{lmn}(\boldsymbol{k}\Delta_L) = ((2l+1)(2m+1)(2n+1))^{1/2} \times j_l\left(\frac{k_x\Delta_L}{2}\right) j_m\left(\frac{k_y\Delta_L}{2}\right) j_n\left(\frac{k_z\Delta_L}{2}\right). \qquad (2)$$

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 $\boldsymbol{k}$.

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 σ_L^2/σ^2 against the spherical top-hat filter mass, M, and size of a spherical perturbation, λ, related via $M = \overline{\rho} \times 4/3\pi r^3$, where $\overline{\rho} = \Omega_0 \times 3{H_0}^2/(8\pi G)$, 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 $M_{\max\,\sigma^2}$ and $\lambda_{\max\,\sigma^2}$, respectively.

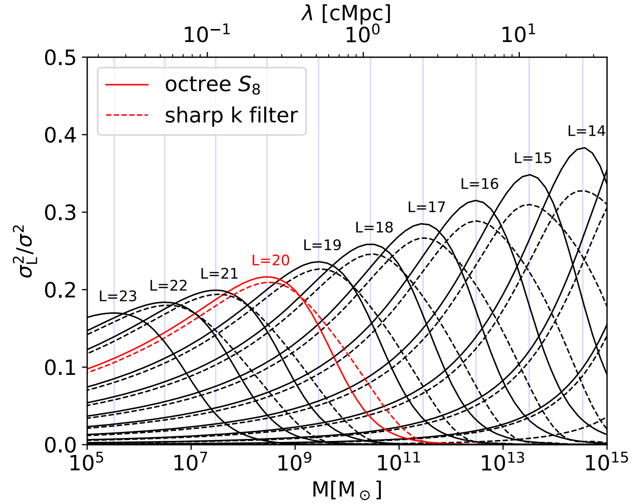

Figure 1. Relative contribution to total the variance of the density field, σ^2, by different octree levels, using the S_8 octree basis functions (solid lines), or a sharp k-space filter (dashed lines), as a function of wavelength, λ, top axis, and corresponding spherical top-hat mass, bottom axis. The scale at which the relative contribution from a given octree level L is maximal defines $\lambda_{\max\,\sigma^2}(L)$ and $M_{\max\,\sigma^2}(L)$, and is indicated by faint vertical lines.

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 2 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 is 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

$$\left\langle P^L_{\mathrm{lin}}(\boldsymbol{k})\right\rangle = P_{\mathrm{lin}}(k)\left(\sum_{l,m,n=0,1} j^2_{lmn}(\boldsymbol{k}\Delta_L)\right). \qquad (3)$$

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

In Fig. 2, we illustrate how the full ΛCDM dimensionless power spectrum is built up from successive higher 'layers' of octree functions by plotting equation (3) for a ΛCDM power spectrum for several values of L. We can see that the ΛCDM 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 cut-off in the spherically averaged power spectrum. We can define an

---

[2] http://virgo.dur.ac.uk/

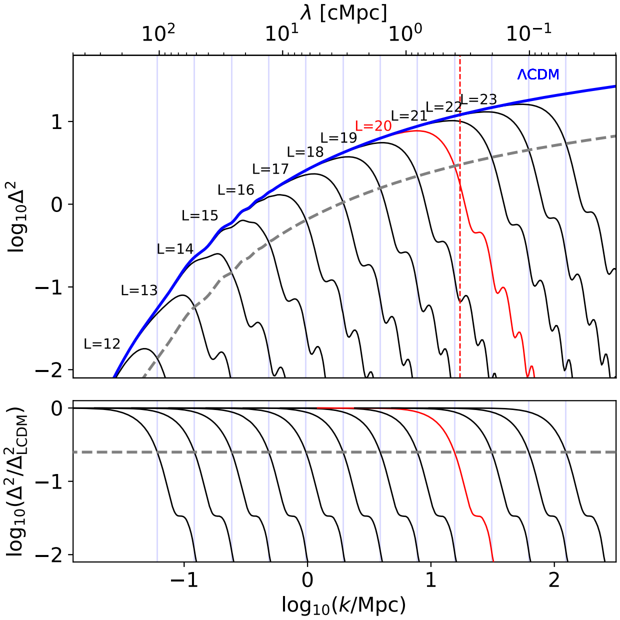

Figure 2. Top panel: the reproduction of the ΛCDM dimensionless power spectrum (blue solid line) as a function of wavenumber, k, bottom axis, and wavelength, λ, top axis, with increasing accuracy. Bottom panel: the ratio of the truncated power spectra and the ΛCDM power spectrum. Individual black lines show the truncated power spectra given by equation (3), successively including the phase information for octree levels up to the indicated level L. Dashed grey lines and the faint vertical lines indicate a reduction of the truncated power spectra by a factor of 4, allowing us to define a cut-off wavenumber, k_{cut}, and cut-off scales, $\lambda_{\mathrm{cut}}(L)$ and $M_{\mathrm{cut}}(L)$ for each octree level. The red solid and dashed lines show the power spectrum truncated at level $L = 20$ and the Nyquist wavenumber of the cells at $L = 20$, respectively.

associated cut-off wavenumber, $k_{\mathrm{cut}}(L)$, by determining where $P^L_{\mathrm{lin}}(k)$ falls by a factor of 2 in amplitude, or 4 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 that relates 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.

Column 3 gives $\lambda_{\max\,\sigma^2}(L)$, the wavelength of a perturbation for which the contribution from level L to the fractional variance σ^2_L/σ^2 in the S_8 parametrization is maximal, as shown in Fig. 1. Column 4 gives $M_{\max\,\sigma^2}$, the equivalent spherical top-hat mass. Column 5 gives $\lambda_{\mathrm{cut}}(L)$, the cut-off scale, i.e. the small-scale limit at which the power due to the truncation at level L falls to a quarter of the value of the full, linear ΛCDM power spectrum, and M_{cut}, in column 6, is the mass of an equivalent spherical top-hat.

While the size of octree cell is of interest, this length scale on its own is not particularly revealing. This is because, as described above, the physical length scales affected by the octree functions result from a combination both of the cell size and the functional forms of the octree functions themselves. In general, basis choices that use high-order polynomials affect smaller physical length scales in units of the octree cell size.

Table 1. Correspondence between octree levels and physical scales.

L	Δ_{cell} (cMpc)	$\lambda_{\max\,\sigma^2}$ (cMpc)	$\log_{10}\left(\frac{M_{\max\,\sigma^2}}{\mathrm{M}_\odot}\right)$	λ_{cut} (cMpc)	$\log_{10}\left(\frac{M_{\mathrm{cut}}}{\mathrm{M}_\odot}\right)$
23	0.065	0.025	5.5	0.051	6.4
22	0.13	0.053	6.5	0.101	7.3
21	0.26	0.11	7.5	0.202	8.2
20	0.52	0.24	8.5	0.405	9.1
19	1.04	0.52	9.5	0.81	10.0
18	2.08	1.11	10.5	1.62	10.9
17	4.17	2.41	11.5	3.24	11.8
16	8.33	5.27	12.5	6.48	12.7
15	16.67	11.6	13.5	12.97	13.6
14	33.33	25.6	14.5	25.94	14.6
13	66.67	56.5	15.6	51.88	15.5
12	133.33	122.9	16.6	103.78	16.4

Note. L: octree level; Δ_{cell}: size of the octree cell at level L; $\lambda_{\max\,\sigma^2}$: characteristic wavelength, where the fractional contribution to the variance from level L is maximal; $M_{\max\,\sigma^2}$: mass of a mean-density spherical top-hat of diameter $\lambda_{\max\,\sigma^2}$; λ_{cut}: wavelength at which power is reduced by four when power at levels L and above are set to zero; M_{cut}: mass of a mean-density spherical top-hat of diameter λ_{cut}.

As shown in Fig. 1, for most scales of interest, the contribution from any one level to the variance is less than 1/3, and any object contains phase information from multiple levels. Comparing the red lines in Figs 1 and 2, we see that, while the octree functions at $L = 20$ contribute most at $\lambda_{\max\,\sigma^2} = 0.24$ cMpc, the power already falls to 1/4 at $\lambda_{\mathrm{cut}} = 0.41$ 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 $\lambda_{\max\,\sigma^2}$ thus locates the peak contribution from a single level, for the effect of variations at and above a certain scale on structure formation, λ_{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' realization 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 that is the phase information used for the EAGLE project flagship 100^3 cMpc^3 volume (Schaye et al. 2015, hereafter called 'Reference'). 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 12th 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).

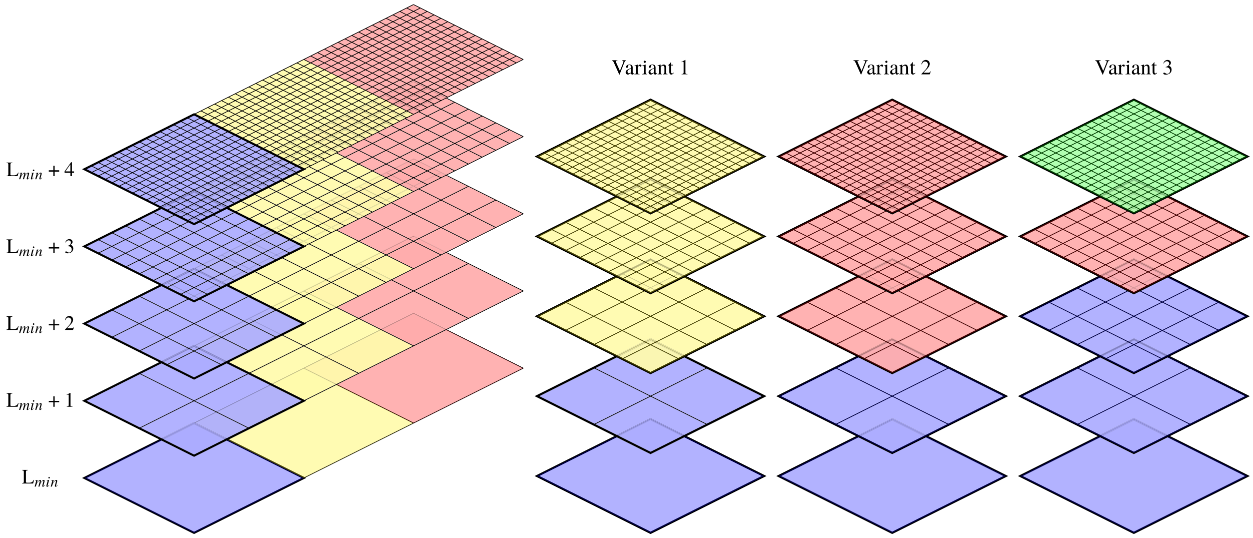

Figure 3. Illustration of displacements in the PANPHASIA white noise field. In all panels the region of PANPHASIA represented in the simulation volume is indicated by thick lines. In the left-hand panel, adjacent regions of PANPHASIA displaced by 1 and 2 in one dimension are shown in red and yellow, respectively. The second and third panels (variants 1 and 2) illustrate shifts at level $L_{\mathrm{min}+2}$ by 1 and 2, respectively. The fourth panel, variant 3, represents a shift by 1 at level $L_{\mathrm{min}+3}$ and another arbitrary shift at level $L_{\mathrm{min}+4}$. Any combination of shifts is possible, and all integer shifts result in independent white noise fields on the respective levels.

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-hand panel shows three adjacent subvolumes of PANPHASIA, in blue, yellow, and red, respectively. Each subvolume 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_{min} to L_{max}, variants 1 and 2 differ from the Reference simulation, and from each other, at levels $L_{\mathrm{min}+2}$ and above. Variant 3 shares levels L_{min} to $L_{\mathrm{min}+2}$ with the Reference simulation, levels $L_{\mathrm{min}+3}$ with the variant 2, and differs from all other simulations at level $L_{\mathrm{min}+4}$.

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 variants 1 and 2, share the same amount of phase information (L_{min} to $L_{\mathrm{min}+2}$), 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 $L_{\mathrm{min}+4}$, the fact that they differ at $L_{\mathrm{min}+3}$ 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 differs from the phase information of the Reference simulation from level L to L_{max} as variants VL, and identify individual volumes that employ a shift by i from level L to L_{max} as VL$_i$. For example, V16 is the set of variants that differ from the Reference simulation at level $L = 16$ and above, and V18$_5$ 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 $M - 1$ and by j from level M to level L_{max}, where $L < M \leq L_{\mathrm{max}}$, are labelled VL$_i$/$M_j$, etc. Hence, V18$_2$/21$_5$ is the individual variant that employs a shift by 2 at levels 18–20, 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 $h = 0.6777$, $n_{\mathrm{s}} = 0.9611$, $\sigma_8 = 0.8288$, $\Omega_0 = 0.307$, $\Omega_{\mathrm{b}} = 0.0483$, and $\Omega_\Lambda = 0.693$. They are set up in a volume of 100^3 cMpc^3 with glass initial conditions, using $N = 376^3 (5.3 \times 10^7)$ particles, giving a particle mass of 7.4×10^8 M$_\odot$, a mean interparticle separation of 266 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. To minimize interpolation errors, we used a 1536^3 Fourier mesh with quadratic interpolation to generated all of the initial conditions. The Nyquist frequency of this Fourier mesh is a factor of 4 smaller than the theoretical particle Nyquist frequency. We set all Fourier modes to zero if the magnitude of their wave vector 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, '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 friends-of-friends (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 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 white noise fields (WNFs) 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-hand 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-hand panels appear almost identical. For variants at level 20, most haloes in the right-hand 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-hand column, all haloes appear visibly displaced, and many low-mass haloes can no longer be matched by eye, while in the left-hand 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 the top panel of Fig. 5, we show the matter power spectrum, $P(k)$, of our simulations measured at $z = 0$. We define

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

as the volume-averaged power spectrum, where δ_k is given by the three-dimensional Fourier transform of the density perturbation field, $\delta(\boldsymbol{x}) = \rho(\boldsymbol{x})/\langle\rho\rangle - 1$, 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 times. 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 13 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.

In the bottom panel, for $k > 1\ h\ \mathrm{Mpc}^{-1}$, we show the cross-power spectrum of the density field of the Reference simulation and each of the variant simulations, in addition to the (auto)power spectrum of the Reference simulation.

It can be seen that, for small variations (e.g. for variations at level 23), all variants have a nearly identical power spectra, and follow all the peculiar features of the Reference simulation. Likewise, at levels 22 and 23, the cross-correlation between the variant and the Reference simulation is nearly identical to the autocorrelation 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 in the autocorrelation, 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.

Differences in the cross-correlation are apparent from level 21 down, and first noticeable on small scales. As expected from Fig. 2, the cross-correlation drops below the autocorrelation significantly above the cut-off scale (or below the cut-off wavenumber).

For larger variations, it can also be seen that the average of the power spectra 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.

All power spectra were computed on a 376^3 mesh created with triangular-shaped-cloud interpolation of the particle data with compensation for the window function, using the NBODYKIT package (Hand et al. 2018). For the cross-correlation, the k-space binning was adapted for each level.

3.3 Halo abundance

In Fig. 6, we compare the halo mass functions in terms of $M_{200,\mathrm{crit}}$ of our simulations to the analytic mass function of Jenkins et al. (2001), calculated for the same cosmology. Because of the small sample size, the Reference simulation (thick black solid line), measured in a single, $100^3\ \mathrm{Mpc}^3$ 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 $L = 13$ to 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 ($\lambda_{\mathrm{cut}} = 13$ cMpc) 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 for each set is greater at higher mass. This may appear counterintuitive,

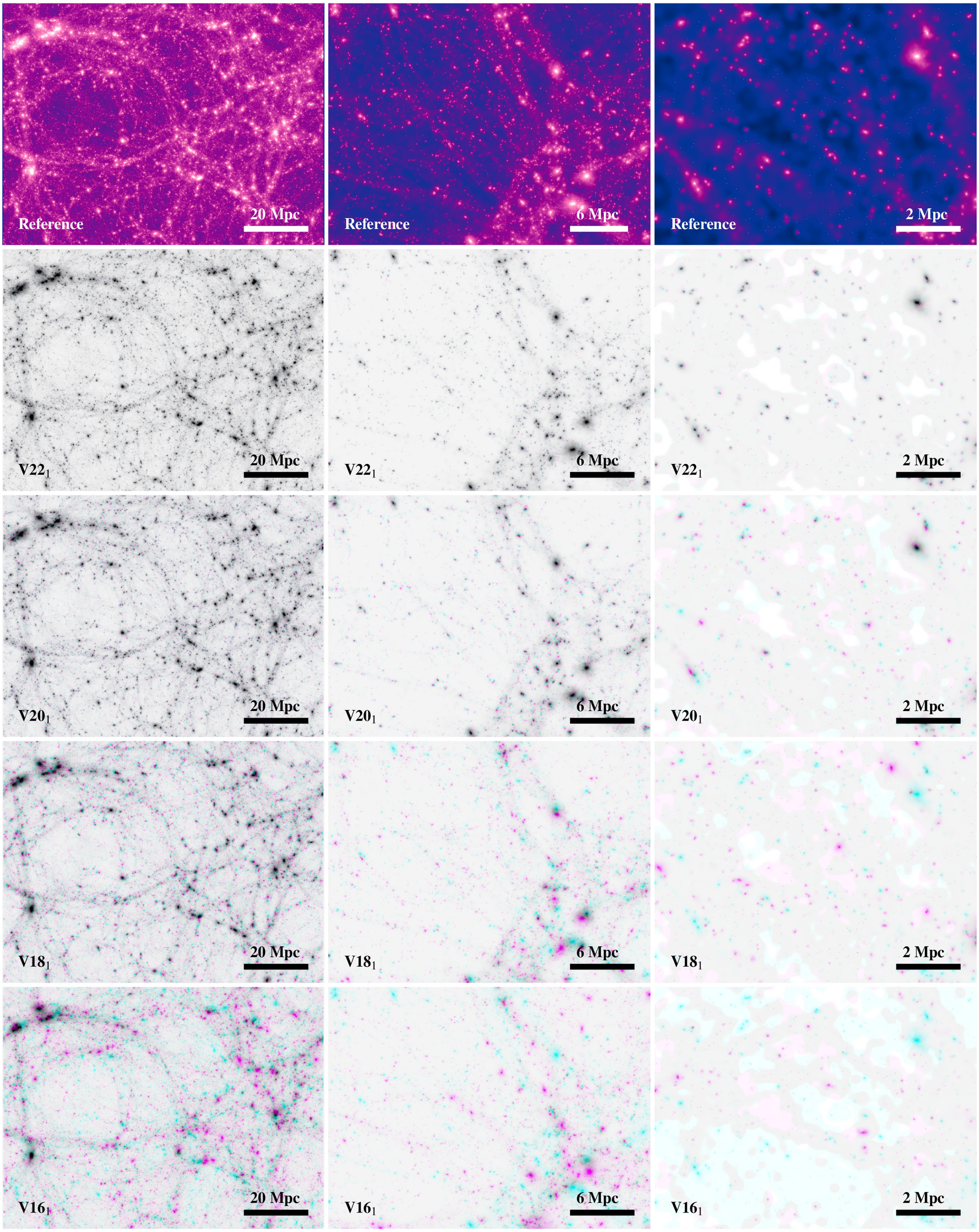

Figure 4. Dark matter density at $z = 0$. From left to right, columns show boxes of width and depth 100, 33, and 10 Mpc, respectively. The top row shows the simulation with the Reference phase information, while the following rows show the effect of randomizing phase information at and above levels 22, 20, and 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, and purple indicates higher density in the variant simulations.

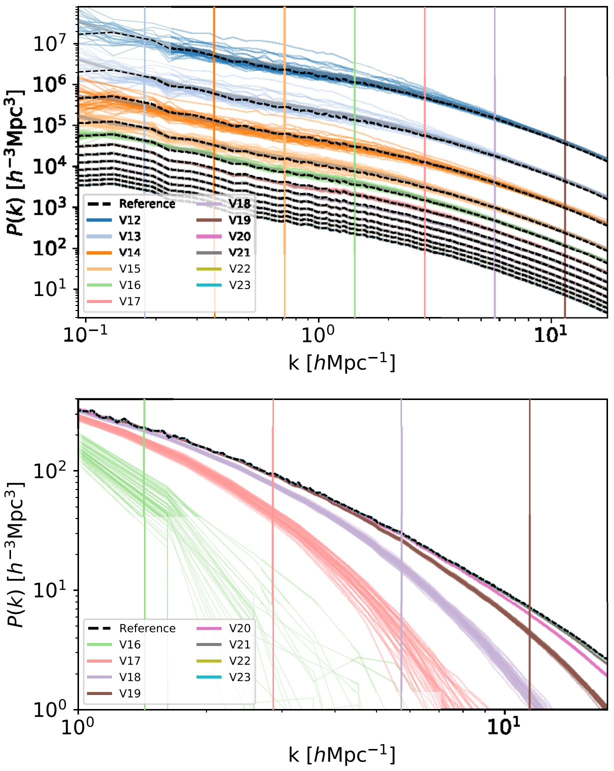

Figure 5. Top: matter power spectra at $z = 0$, for the Reference simulation (black dashed) and 39 variants at each level from 12 to 23. For visual clarity, the power spectra of the variant simulations are offset, and the result of the Reference simulation is repeated with each set. Solid grey lines show the power spectrum averaged over all variants at each level. Bottom: power spectra of the density cross-correlation between the Reference simulation and the 39 variant simulations at each level from 16 to 23. The dotted black line shows the (auto)power spectrum of the Reference simulation. On both panels, vertical lines show the scale of $k_{\rm cut}$, values for $L > 19$ lie outside the range.

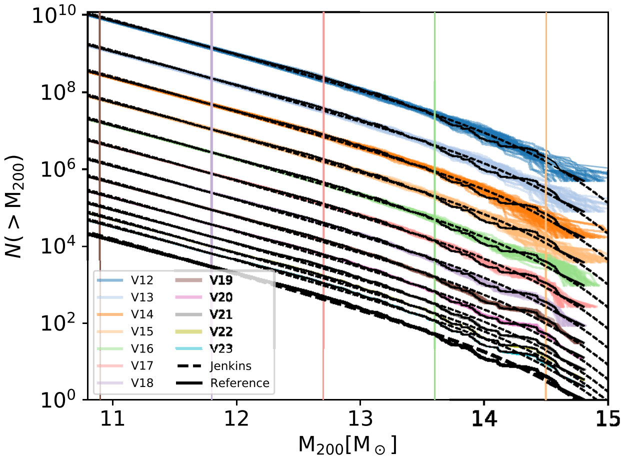

Figure 6. Halo mass function ($M_{200,\rm crit}$) at $z = 0$, for the Reference simulation (thick black solid), and 39 variants at each level from 12 to 23. Vertical coloured lines show the value of $M_{\rm cut}$, the mass of a spherical top-hat of mean density and diameter $\lambda_{\rm cut}$. Also shown (thick black dashed) is the halo mass function of Jenkins et al. (2001), calculated for the same cosmology. For visual clarity, the sets of halo mass functions of the variant simulations are offset, and the mass functions of the Reference simulation, and of Jenkins et al. (2001) are repeated with each set, as thin solid and dashed lines, respectively.

Table 2. Median and standard deviation in the number haloes of a given mass across the simulations of a given level.

Level	$N(10^{11})$	$N(10^{12})$	$N(10^{13})$	$N(10^{14})$
Ref	21 689	2602	343	21
23	21 723 ± 48	2613 ± 9	343 ± 1	21 ± 0
22	21 683 ± 51	2597 ± 6	345 ± 2	21 ± 0
21	21 676 ± 78	2613 ± 16	341 ± 3	21 ± 0
20	21 704 ± 81	2629 ± 18	344 ± 6	21 ± 1
19	21 790 ± 63	2698 ± 28	347 ± 8	21 ± 1
18	21 833 ± 118	2662 ± 46	356 ± 12	23 ± 2
17	21 825 ± 257	2703 ± 28	328 ± 8	28 ± 2
16	21 711 ± 214	2674 ± 66	327 ± 22	27 ± 3
15	21 403 ± 266	2655 ± 80	318 ± 13	32 ± 5
14	21 705 ± 292	2732 ± 100	322 ± 16	29 ± 5
13	21 554 ± 184	2677 ± 87	318 ± 25	29 ± 5
12	21 523 ± 193	2682 ± 89	318 ± 25	30 ± 6

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, σ_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^{12} M$_\odot$ is ~2700–52^2, 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 $L = 19$ lead to the formation of independent 10^{11} M$_\odot$ haloes, their population is so well sampled in the 100 Mpc volume that their number has less than a 1 per cent 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 level 12 or 13, we can see that the Reference simulation contains slightly fewer than the expected 30 ± 5 haloes of mass 10^{14} M$_\odot$.

4 INDIVIDUAL 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

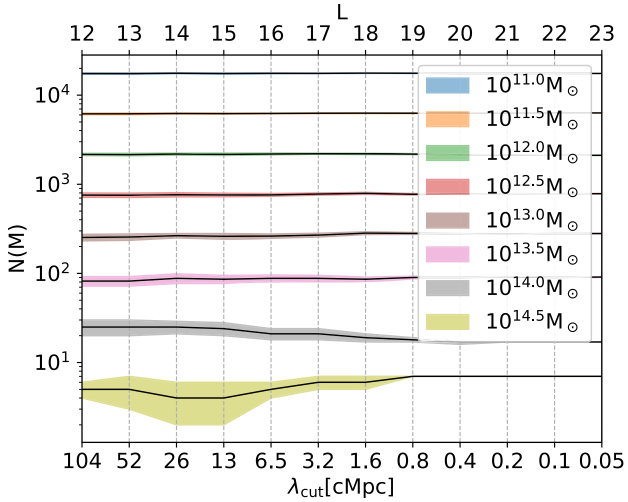

Figure 7. Median and standard deviation in the number of haloes for different masses, as a function of the level at which variations are introduced in the initial conditions at $z = 0$, L, and the corresponding wavelength, $\lambda_{\rm cut}$. The Reference simulation has the expected number of haloes up to $\sim 10^{13.5}$ M$_\odot$, but lies outside 1σ for both $\sim 10^{14}$ and $\sim 10^{14.5}$ M$_\odot$ haloes.

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 3 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), 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 almost no phase information, the matching rate only falls to $\sim$15 per cent. Two haloes matched under these conditions have, by coincidence, grown to similar mass from overlapping Lagrangian volumes, without their simulations sharing any relevant information. Although matched, these are not physically the same halo.

4.1 Halo identity across simulations

The possibility that two similar haloes can exist in two volumes that share no phase information leaves the tantalizing 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, in each panel, we show the multiple-matching rate, f_N, as a function of mass, for 1–39 possible matches between the Reference simulation and all $N = 39$ variants at each level. Shades of red denote the fraction of haloes with 1–19 matches; shades of blue correspond to 20–39 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 per cent, 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* tends to zero for low-mass haloes and low-level variations, and to unity for high masses and high-level variations.

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: we say that a halo *exists* at and above level $L_{\rm E}$ if it can be matched to more than half of the simulations that randomly vary the initial density fields at levels above $L_{\rm E}$. In other words, a halo exists at a scale $L_{\rm E}$ because, at this scale, the initial density field contains the necessary information for the formation of this particular halo.

In Fig. 10, we show the fraction of unique haloes that can be matched to more than half of the variants, $f_{\rm match,1/2}$, as a function of level, L, and of the corresponding cut-off wavelength, $\lambda_{\rm cut}(L)$. Circles show the results measured in our simulations, dashed lines of corresponding colours show two-parameter logistic fits of the form

$$f_{\rm match,1/2}(\lambda_{\rm cut}, M) = \frac{1}{1 + {\rm e}^{-a_M(\lambda_{\rm cut} - \lambda_{{\rm E},M})}}, \qquad (4)$$

where a_M and $\lambda_{{\rm E},M}$ are free parameters fit separately at each mass, over the domain indicated by the extent of the dashed lines. Crosses indicate the radius, $r_{\rm halo}(M)$, of a mean-density sphere of mass M.

For each halo mass, we find a similar behaviour, with values of a_M in the range 8–10. Moreover, we find linear scaling of $\lambda_{{\rm E},M} \propto M^{\sim 1/3} \propto r_{\rm halo}(M)$. This reflects the fact that the length scale on which the initial density field needs to be defined for unique haloes to exist, is, on average, closely related to the size of the Lagrangian volume that will collapse into the halo.

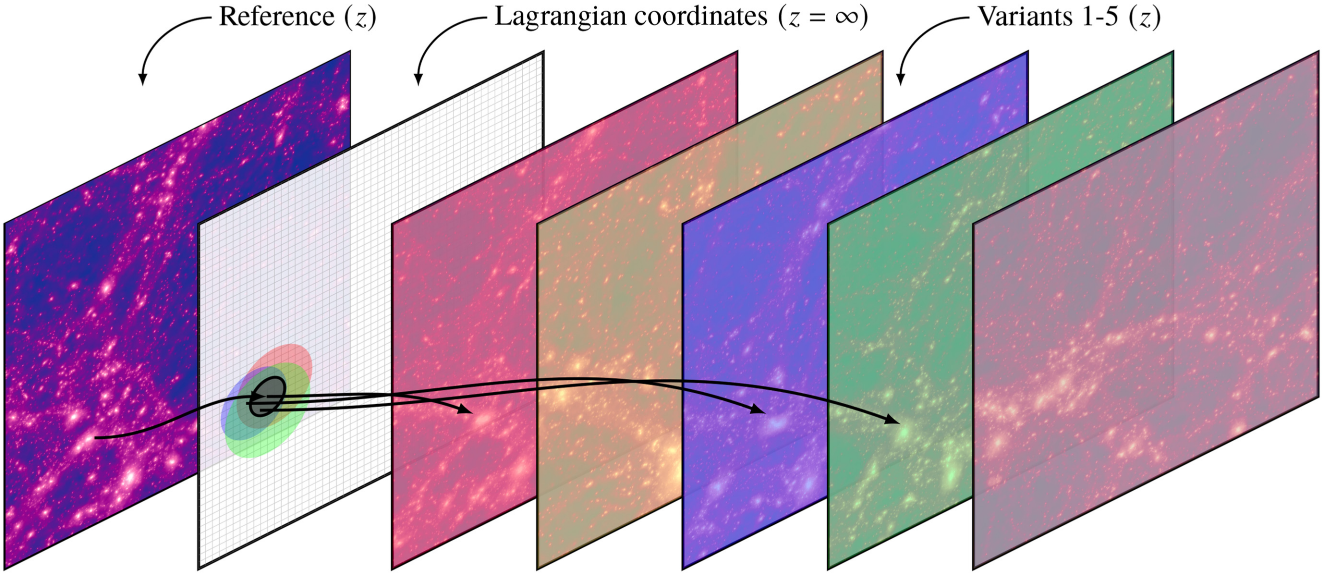

Figure 8. Matching of haloes using particles with identical coordinates at $z = \infty$. The origin of the most bound particles in a given halo of the Reference simulation at redshift z ('Reference particles') define a Lagrangian volume (denoted by the grey circle) to which all particles in haloes in the variant simulations are compared. In this example, variants 1 (red), 3 (blue), and 4 (green) each contain a halo that includes the majority of Reference particles, and whose mass at redshift z is within a factor of 3 of the halo in the Reference simulation. By contrast, variants 2 and 5 contain no matching haloes.

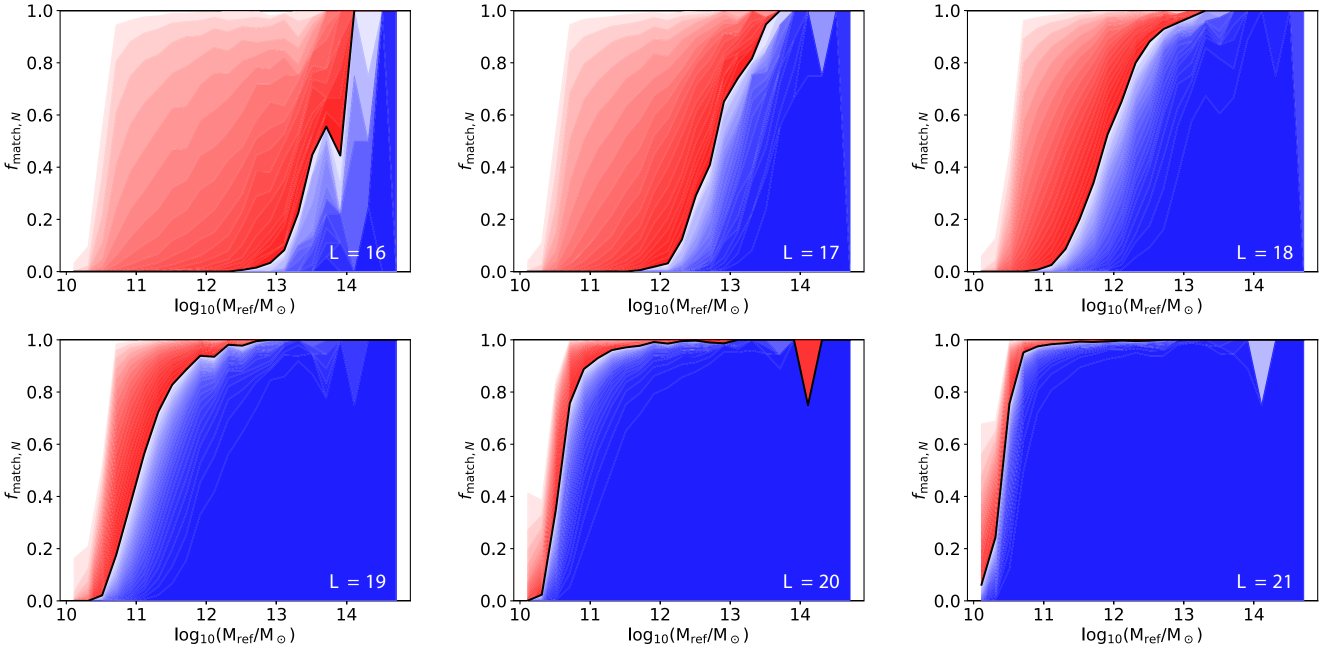

Figure 9. Fraction of haloes with multiple matches, $f_{\mathrm{match},N}$, between the Reference simulation and the 39 variant simulations at each level from 16 to 21, as a function of halo mass in the Reference simulation. Shades of red and blue indicate 1–19 and 21–39 matches, respectively, while 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.

Accounting for this self-similar behaviour, the solid lines in Fig. 10 show a global fit to equation (4), with $a_M = a = 9$, and

$$\log_{10}(\lambda_{\mathrm{E},M}/\mathrm{cMpc}) = 1/3\ \log_{10}(M/\mathrm{M}_\odot) - 3.79, \tag{5}$$

or equivalently,

$$\lambda_{\mathrm{E},M} = 0.88 \times r_{\mathrm{halo}}(M). \tag{6}$$

Dashed lines give a slightly 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 two free parameters, a and the coefficient in equation (6) that determines $\lambda_{\mathrm{E},M}$.

The phase information of the initial density field has to be defined at least down to $\lambda_{\mathrm{cut}} = \lambda_{\mathrm{E},M}$ in order to not only create a similar

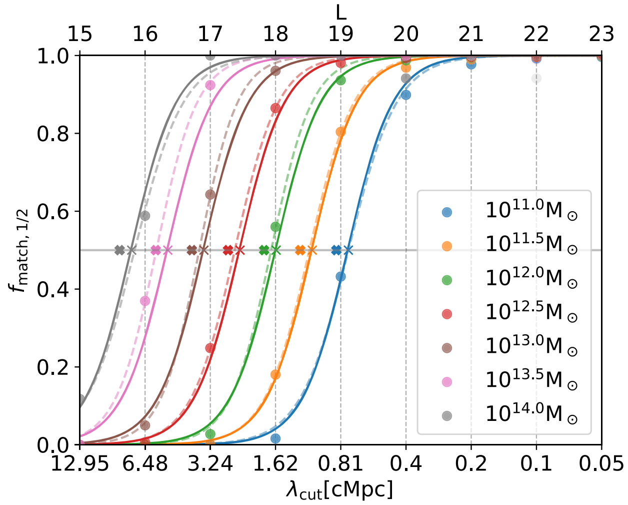

Figure 10. $f_{\rm match,1/2}$, the fraction of haloes at $z = 0$ that are matched more than half of the time across variations at a given cut-off wavelength ($\lambda_{\rm cut}$, top axis), and level (L, bottom axis), for different masses. The scale at which $f_{\rm match,1/2}$ equals 1/2 defines $L_{{\rm E},M}$ and $\lambda_{{\rm E},M}$, i.e. the level and wavelength at which there is sufficient information in the primordial white noise field for the formation of unique $z = 0$ haloes of a given mass, M. Circles show results measured in the simulations, dashed lines are individual fits to equation (4) at each mass, and solid lines are fits of the same equation to the entire data set. For comparison, crosses show $r_{\rm halo}(M)$, i.e. the radius of a mean-density sphere of mass M. It can be seen that $\lambda_{{\rm E},M}$ scales with mass proportional to $r_{\rm halo}$, indicating a self-similar behaviour.

number of haloes of a given mass at $z = 0$, but also to determine the formation of unique haloes of mass M. We find that this scale is ∼88 per cent of $r_{\rm halo}(M)$, the radius of the Lagrangian volume enclosing a mean-density sphere of mass M, and ∼44 per cent of the wavelength, $\lambda_{\rm cut}$, of a perturbation expected to collapse into a halo of mass $M_{\rm cut}$. As examples, we find that, at $z = 0$, unique haloes of 10^{12} and 10^{14} M$_\odot$ *exist* when the initial density field is defined down to cut-off wavelengths of 1.6 and 7.5 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 realizations. 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 realization of that halo. If we consider that the possible random variations on scales below $L_{\rm E}$ 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 realizations of a halo to the Reference simulation. Instead, we compute a median value from all realizations at a given level, and analyse the scatter among individual samples.

5.1 Mass

Fig. 11 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 18–23. 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

$$\sigma^*_M = \sigma(\log_{10} M). \tag{7}$$

The red bands on each panel of Fig. 11 indicate σ^*_M, in bins equally spaced in $\log(\langle M\rangle)$. 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 18. For haloes of a given mass, the mass varies more for lower level variations.

This behaviour is summarized across masses and levels in Fig. 12, which shows $\sigma^*_{\rm M}$ as a function of level and $\lambda_{\rm cut}$. 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

$$\sigma^*_M = a_M \lambda_{\rm cut} + b_M, \tag{8}$$

which are shown in Fig. 12 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 $L_{\rm E}(M)$, given by equation (5), data points below $L_{\rm E}(M)$ 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 $a = a_M = 1.15$ for all mass bins, and a mass dependence for $b_M = 4.6 - 0.445\,\log_{10}(M/{\rm M}_\odot)$, i.e.

$$\sigma^*_M = 4.6 + 1.15\lambda_{\rm cut} - 0.445\,\log_{10}(M/{\rm M}_\odot). \tag{9}$$

Fits to this universal relation are shown by solid lines in Fig. 12.

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, $L_{\rm E}$, we find that the scatter in halo mass typically reaches $\sigma^*_M \sim 0.23$, or a factor of ∼1.7 in mass before haloes drop out of existence. Scatters in mass of ∼10 per cent ($\sigma^*_M = 0.041$) or ∼1 per cent ($\sigma^*_M = 0.0043$) are found for variations at 2 and 5 levels above the existence level $L_{\rm E}$, respectively.

---

[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, $\sigma^*_x = \sigma\,(\log_{10}(x))$.

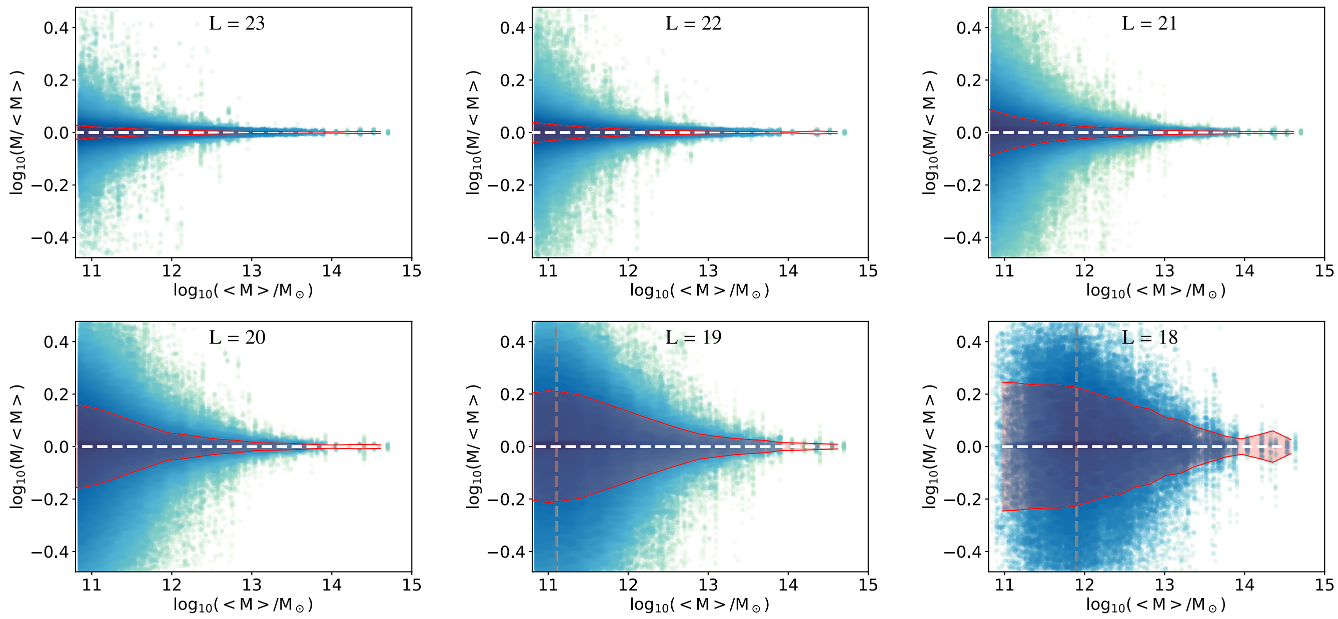

Figure 11. Mass ratio between the mass of individual, matched central haloes in all 39 variant simulations at a given level, and their average mass across all 40 simulations, as a function of average mass. The red band shows the 1σ scatter. Vertical dashed lines on panels at $L = 18$ and 19 indicate the minimum mass of haloes that *exist* at these level.

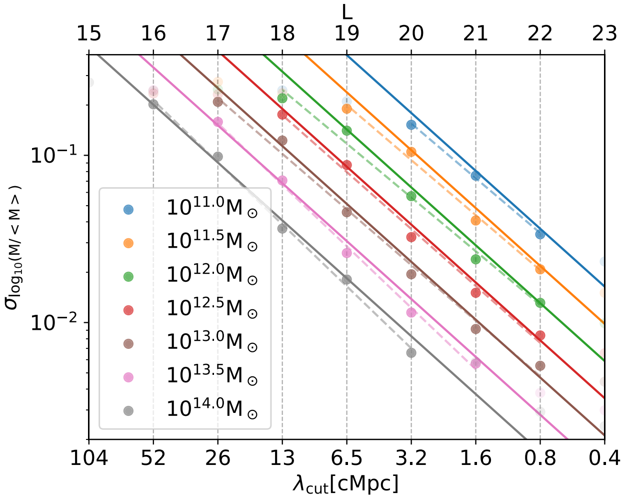

Figure 12. Median scatter in log halo mass, σ_M^* for matched haloes of different mass (indicated by different colours), as a function of level of randomization in the initial density field (L, upper x-axis), and corresponding wavelength λ_{cut}. Only bold data points are included in the fits. Dashed lines show independent power-law fits for each mass bin, solid lines show a universal power-law fit for all masses.

5.2 Concentration

In addition to the mass, a second parameter is required to characterize a ΛCDM halo whose density is described by a Navarro–Frenk–White (NFW) profile (Navarro, Eke & Frenk 1996),

$$\rho = \frac{\rho_{\mathrm{s}}}{(r/r_{\mathrm{s}})(1+r/r_{\mathrm{s}})^2}, \tag{10}$$

where r_{s} and ρ_{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

$$\delta_{\mathrm{c}} = \frac{200}{3}\frac{c^3}{\ln(1+c) - c/(1+c)}, \tag{11}$$

where δ_{c} is the characteristic overdensity.

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

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

For an NFW halo, the characteristic overdensity, δ_{c}, of equation (11) is related to the overdensity inside $r_{\max}$, δ_v, via the relation

$$\delta_{\mathrm{c}} = \frac{\rho_{\mathrm{s}}}{\rho_{\mathrm{crit}}} = 7.213\,\delta_v.$$

We can thus calculate the concentration parameter, c, of each halo from the measured values of $r_{\max}$ and $v_{\max}$ by computing δ_v and δ_{c}, and solving equation (11) for c.

Because stripping of satellites affects $r_{\max}$ disproportionately relative to $v_{\max}$, equation (11) 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 $r_{\max}$. 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 7.4×10^{11} $\mathrm{M}_\odot$.

Neto et al. (2007) showed that concentration parameters for relaxed haloes of a given mass are approximately lognormal distributed, with $\sigma_c^* \sim 0.11$ at 10^{12} $\mathrm{M}_\odot$. Fig. 13 shows the relative scatter in concentration of matched central haloes, as a function of median mass, and for variations at levels 18–23. The red band shows σ_c^*.

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^{13} $\mathrm{M}_\odot$ haloes at $L = 22$

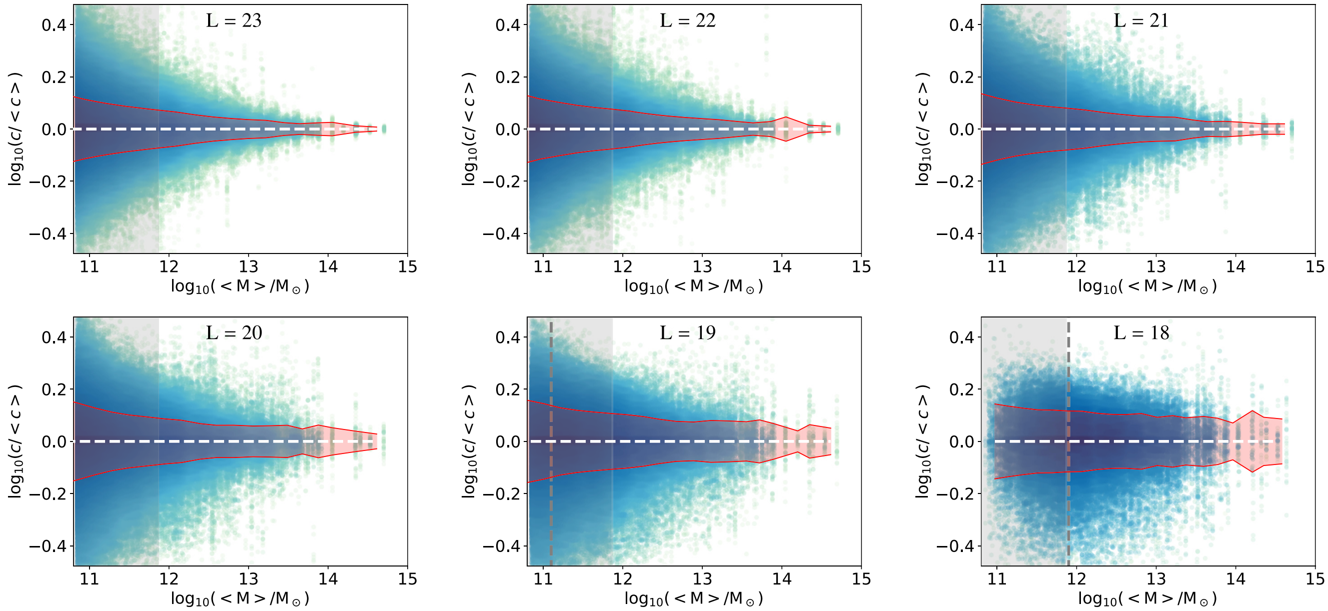

Figure 13. Ratio between concentration of matched central haloes in the variant simulations and the median concentration for each halo, as a function of median halo mass, for different levels (indicated on each panel). The red regions show the standard deviation among variants. Grey regions denote haloes with fewer than 1000 particles, the convergence limit determined by Neto et al. (2007); vertical dashed lines on panels at $L = 18$ and 19 indicate the minimum mass of haloes that *exist* at these levels.

is 8 per cent, while the scatter in mass is <1 per cent for the same variation. This reflects the fact that the scale that determines the concentration of a halo, $r_{\rm max}$, is much smaller than its total size, and we find a much greater scatter in $r_{\rm s}$ 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 ΛCDM 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. 14 shows σ_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 ∼0.1, σ_c^* appears to scale linearly with $\lambda_{\rm cut}$, 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. 15 shows the relative displacement, Δ_r, of individual, matched haloes at $z = 0$ from their median position, as a function of their median mass, and for variations at levels 18–23. Unlike in Fig. 11, the range on each panel is adjusted to include 99 per cent of points in every case.

We calculate the scatter in position as

$$\sigma_r = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}, \tag{12}$$

where $r(x, y, z)$ are the Cartesian coordinates across the variants of the halo, and σ_x, σ_y, and σ_z are the 1D dispersions in position.

In Fig. 15, σ_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 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.

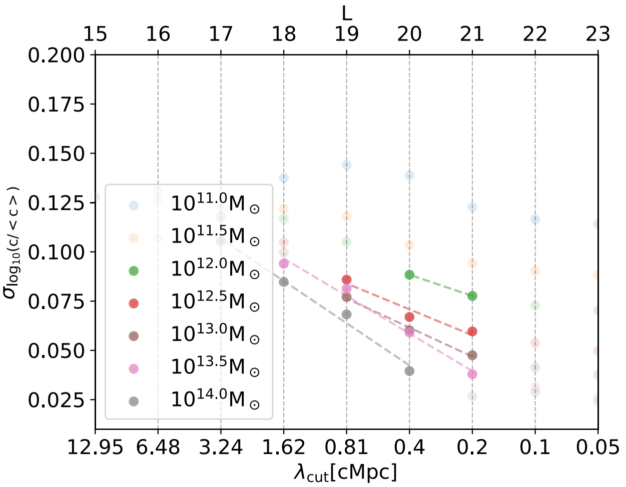

Figure 14. Median standard deviation of halo concentrations for haloes of different mass (indicated by different colours), and as a function of level of randomization in the initial density field (L, upper x-axis), and corresponding cut-off wavelength $\lambda_{\rm cut}$. Only bold data points are included in the fits. Dashed lines show independent power-law fits for each mass bin.

The dependence of average displacement on halo mass and level is illustrated in Fig. 16. Different coloured symbols show the scatter

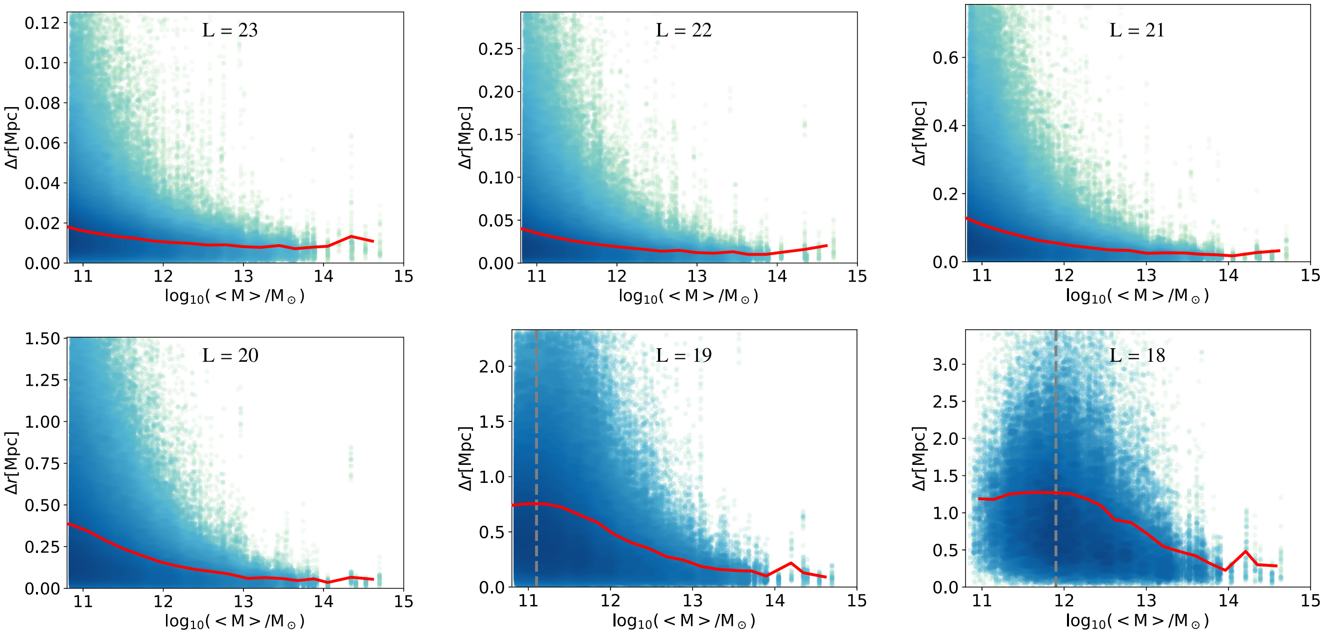

Figure 15. Displacement of matched central haloes in the variant simulation relative to the median position for each halo, as a function of median halo mass, for different levels (indicated in each panel). The red lines show one standard deviation in position among variants. Vertical dashed lines in 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.

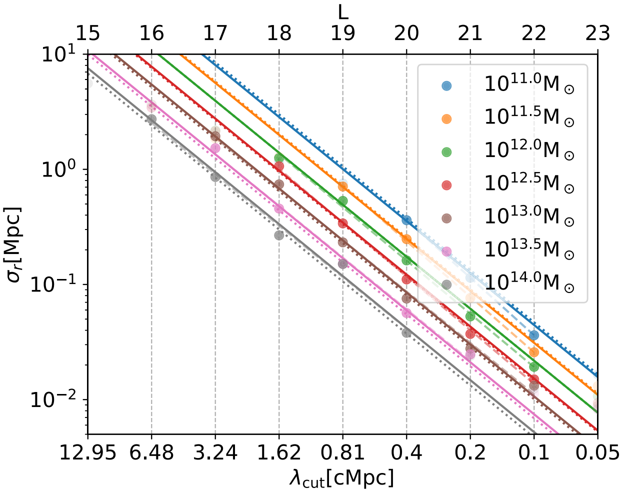

Figure 16. Scatter of halo positions, σ_r for central haloes of different mass (indicated by different colours), and as a function of level of randomization in the initial density field (L, upper x-axis), and corresponding cut-off wavelength ($\lambda_{\rm cut}$, lower x-axis). Only bold data points 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. Dotted lines show a universal fit with the scaling proportional to $r_{\rm halo}$.

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 $\lambda_{\rm cut}$ of the form

$$\log_{10}(\sigma_r/{\rm cMpc}) = a_M \log_{10}(\lambda_{\rm cut}/{\rm cMpc}) + b_M. \qquad (13)$$

Dashed lines Fig. 16 show separate fits at each mass, when the same limits to the domain as in Section 5.1 are applied. As in equations (4) and (8), we find a self-similar behaviour, with a universal slope of $a =$ $a_M = 1.5$, and a mass dependence of $b_M = 3.55 - 0.31\log_{10}(M/{\rm M}_\odot)$. Solid lines show fits with these parameters of a and b_M at every level. When we further restrict b_M to scale proportional to the size of the Lagrangian region, $r_{\rm halo}(M)$, we find $b_M = 3.83 - 1/3\log_{10}(M/{\rm M}_\odot)$. This relation is shown by the dotted lines in Fig. 16.

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 σ_r to those of σ^*_M, we find a stronger scale dependence for σ_r, and a stronger mass dependence for σ^*_M.

5.4 Velocity

By analogy to Fig. 15, Fig. 17 shows the scatter in velocity of individual, matched haloes at $z = 0$, as a function of their median mass, for variations at levels 18–23. The range in each panel is adjusted to include 99 per cent of points in every case.

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

$$\sigma_v = \sqrt{\sigma(v_x)^2 + \sigma(v_y)^2 + \sigma(v_z)^2}, \qquad (14)$$

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. 18. Different coloured symbols show the scatter measured from our simulations for haloes of different mass, from $10^{10.5}$ M$_\odot$ (blue) to $10^{10.5}$ M$_\odot$ (grey). We find the familiar power-law behaviour of equations (8) and (13), and parametrize the

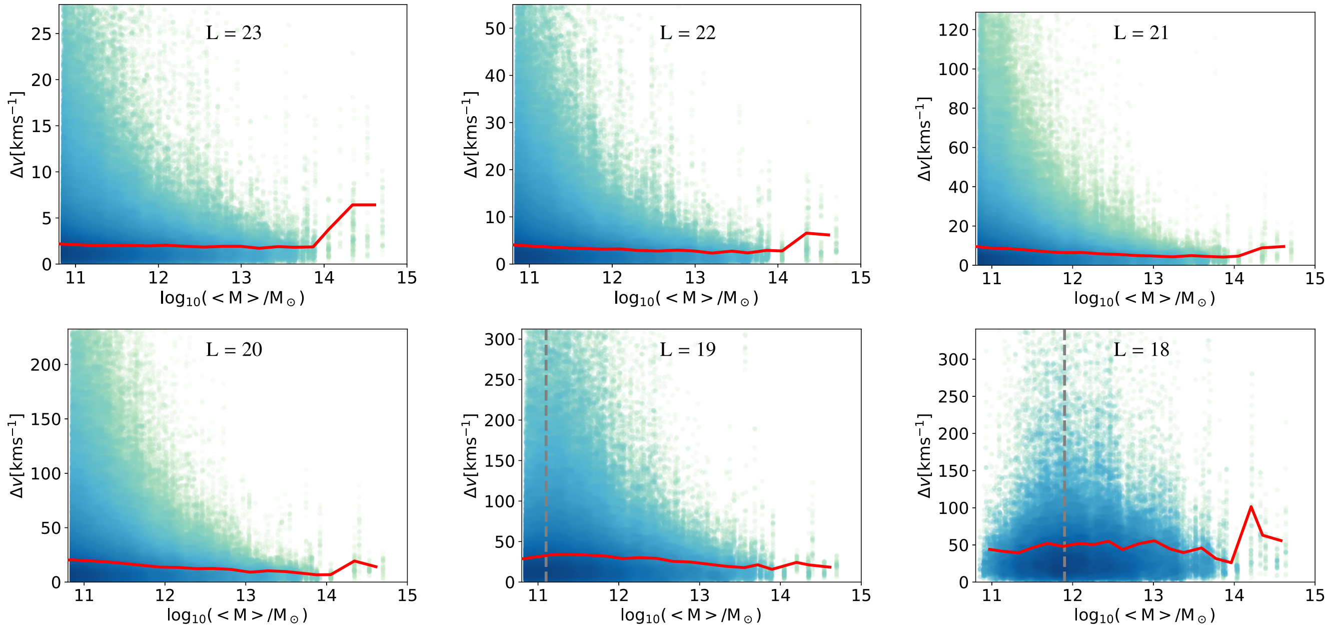

Figure 17. Velocity offset of matched central haloes in the variant simulation relative to the average position, as a function of average mass (shown on the x-axis), and for different levels (indicated in each panel). The red lines show one standard deviation in position among variants. Vertical dashed lines on panels at $L = 18$ and 19 indicate the minimum mass of haloes that *exist* at these level. In each panel, the y-axis extends to the 99th percentile for all matched haloes.

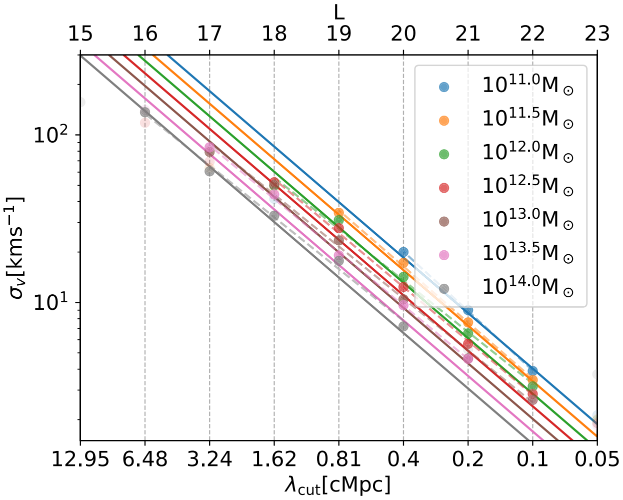

Figure 18. Magnitude of the average velocity difference, relative to the average velocity across variants, for central haloes of different mass (as indicated by colours), and as a function of level of randomization in the initial density field (L, upper x-axis), and corresponding cut-off wavelength (λ_{cut}, lower x-axis).

velocity offset as

$$\log_{10}(\sigma_v/(\mathrm{km\,s^{-1}})) = a_M \log_{10}(\lambda_{\mathrm{cut}}/\mathrm{cMpc}) + b_M. \tag{15}$$

Dashed lines show separate fits at each mass, solid lines show fits assuming a fixed slope at every level. The same cuts to the domain as in Sections 5.1 and 5.3 are applied. Solid lines assume a fit to equation (15) with a universal slope and regular mass dependence, and we find values of $a_M = 1.1$ and a mass dependence of $b_M = 3.35 - 0.15\log_{10}(M/\mathrm{M}_\odot)$.

The velocity dispersion of halo variants increases as the scale of the variation in the initial density field is increased, and also scales inversely with halo mass. Comparing the mass- and scale dependencies of σ_v to those of σ_r, we find an even stronger scale dependence and even weaker mass dependence for σ_v. This can be explained because the velocity of a halo is set not by its internal mass (or even the density of its environment), but by larger scale tidal fields. Haloes respond to a change in environment, but the mass dependence of the velocity on halo mass appears much weaker than for both the mass scatter and displacement.

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 2.2×10^{14} M$_\odot$, whose density map is shown in the top left-hand panel of Fig. 19, 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, Navarro & Frenk (1998), and more recently by Borgani et al. (2002), Kay et al. (2004), Nagai, Vikhlinin & Kravtsov (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 reference simulation. 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.

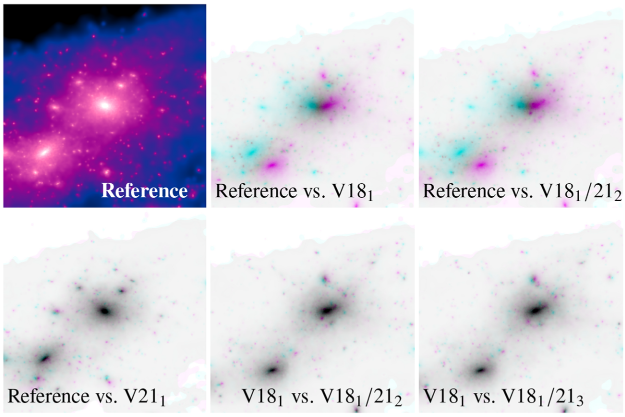

Figure 19. Top left: dark matter density in the most massive cluster in the Reference simulation at $z = 0$. Analogous to Fig. 4, the remaining panels show differences in the projected density in the same Eulerian volume. The top centre, top right-, and bottom left-hand panels show the difference relative to the Reference simulation, for variants V18_1 and V$18_1/21_2$ (top middle and right), and V21_1 (bottom left). As expected, the differences in density relative to the reference simulation are similar in V18_1 and V$18_1/21_2$, which use the same shift on levels $L = 18$–20. By comparison, the difference between V21_1 and the Reference simulation is much smaller. The bottom centre and bottom right-hand panels show the density relative to V18_1 for variants V$18_1/21_2$ and V$18_1/21_3$, respectively. While their differences relative to the Reference simulation are similar in magnitude to that of V18_1, their differences relative to V18_1 are much smaller, and similar in magnitude to that between V21_1 and the Reference simulation.

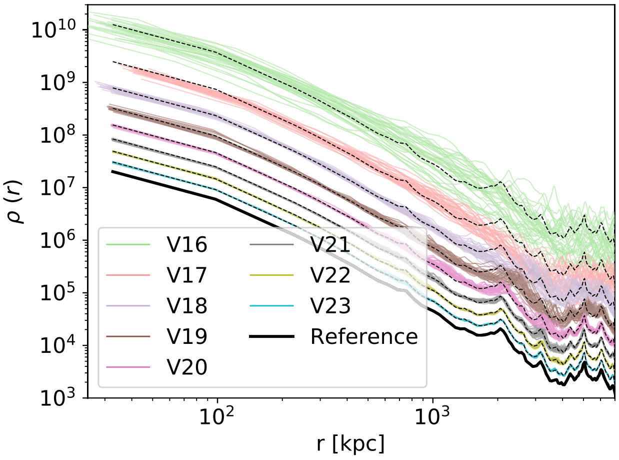

Figure 20. Density profile of the most massive cluster at $z = 0$. The thick black line shows the mass profile in the Reference simulation, while lines of different colour show up to 39 matched haloes at each level from 16 to 23. For clarity, the lines for each level are offset, and the thin black line repeats the Reference simulation with the corresponding offsets.

6.1 Simple variations

In Fig. 20, 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.

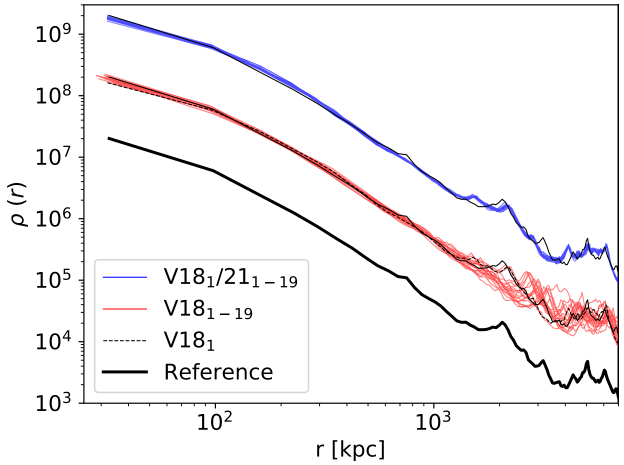

Figure 21. Density profile of the most massive cluster at $z = 0$, similar to Fig. 20. The thick black line shows the mass profile in the Reference simulation. Brown lines show 19 variations at level 18, offset for clarity. The Reference simulation and variant V18_1 are repeated as thin solid and dashed lines black, respectively. Blue lines show 19 additional variations of V18_1 at $L = 21$: V$18_1/20_2$ to V$18_1/21_20$. It can be seen that these result in small perturbations, not of the cluster in the Reference simulation, but of the variation of cluster in V18_1.

At levels $L = 20$, 18, and 16, the scatter in mass is 1 per cent, 10 per cent, and 60 per cent, respectively, while the scatter in concentration is 9 per cent, 20 per cent, and 35 per cent. 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 18–20, and by 1–19 at levels 21 and above. We label this new set of simulations 'V$18_1/21_{[1-19]}$' and note that V$18_1/21_1$ is identical to V18_1.

In Fig. 19, 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-hand panels, it can be seen that compared to the Reference simulation, V18_1 and V$18_1/21_2$ show the same amount of difference, while V18_1, shown on the bottom left, is much more similar to the Reference simulation.

The bottom middle and right-hand panels show the density relative to V18_1 for variants V$18_1/21_2$ and V$18_1/21_3$, respectively. While their differences relative to the reference simulation are similar in magnitude to that of V18_1, their differences relative to V18_1 are much smaller, and similar in magnitude to that between V21_1 and the reference simulation.

Similar to Fig. 20, in Fig. 21, we show the density profiles of the cluster in the 19 variants V18_{1-19} (red lines), and the 19 variants V$18_1/21_{1-19}$ (blue lines), with offsets for clarity. It can be seen that the scatter among the density profiles for $V18_1/21_{1-19}$ is much reduced compared to the scatter among $V18_{1-19}$, and furthermore, that the individual variants of $V18_1/21_{1-19}$ scatter around the variation $V18_1$ (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 localization of our basis allows us to independently vary spatial subvolumes 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 localized, 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.

The precise correspondence between a modification at a certain level, and an effect on a particular scale, depends on our parametrization of the initial density field and our choice of basis functions. As discussed in Section 2.1.1, a modification on one level percolates over a range of wavenumbers and a range of physical scales, both in the initial and the evolved density field. Expressing the modifications in terms of physical scales such as λ_{cut} and $\lambda_{\max\sigma^2}$ alleviates some of the parameter dependence, but we expect that a different implementation will yield slightly different results.

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_{E}, 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 equations (5) and (6), we have parametrized the existence scale as a function of halo mass and the size of its Lagrangian volume.

Beyond the mere existence of haloes, we have also quantified changes in their properties such as mass, concentration, position, and velocity, and find 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} $\mathrm{M}_{\odot}$ 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, Pontzen & Peiris (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 genetic modification. 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

We thank the anonymous reviewer for constructive and insightful suggestions that improved this paper. 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 DMIDAS (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.

DATA AVAILABILITY

The initial conditions for the simulations analysed in this work can be generated using the information described in Appendix A. The simulation data generated as part of this work will be shared on reasonable request to the corresponding author.

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, ApJ, 492, 439

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, MNRAS, 458, 900

Cole S., Kaiser N., 1989, MNRAS, 237, 1127

Colombi S., Szapudi I., Jenkins A., Colberg J., 2000, MNRAS, 313, 711

Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371

Doumler T., Gottlöber S., Hoffman Y., Courtois H., 2013, MNRAS, 430, 912

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

Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160

Hoffman Y., Ribak E., 1991, ApJ, 380, L5

Hoffman Y., Courtois H. M., Tully R. B., 2015, MNRAS, 449, 4494

Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894

Jasche J., Leclercq F., Wandelt B. D., 2015, J. Cosmol. Astropart. Phys., 01, 036

Jenkins A., 2010, MNRAS, 403, 1859

Jenkins A., 2013, MNRAS, 434, 2094

Jenkins A., Booth S., 2013, preprint (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

Lavaux G., Jasche J., 2016, MNRAS, 455, 3169

Linde A., 2005, preprint( )

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., McQuinn M., 2012, ApJ, 760, 4

Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ

Planck Collaboration XVII, 2016, A&A, 594, A17

Planck Collaboration IX, 2020, A&A, 641, A9

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., 2020, Nature, 585, 39

White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

Yepes G., Gottlöber S., Hoffman Y., 2014, New Astron. Rev., 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 × 12 × 12 root cell at level $L = 16$. At level 16 for example the reference phases in the x-direction are taken from the range of 12 cells 31250–31261, and similarly 23438–23449 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 VL_j 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 cMpc, and similarly for y and z:

$$\begin{aligned}\Delta x &= (j \bmod 10),\\ \Delta y &= ((j - \Delta x) \bmod 100)/10,\\ \Delta z &= ((j - 10\Delta y - \Delta x) \bmod 1000)/100.\end{aligned}$$

For example, $\mathrm{V21}_{24}$ is constructed using shifts of $\Delta x = 4$ and $\Delta y = 2$.

For every variant, we shift up to 10 levels at $L_{\rm min}$ 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, $l_{\rm cell} = 100/12/2^{\rm L-16}$ cMpc. In order to shift the phase information by the box size, at each level, a shift by $\Delta N_i = \Delta_i \times 12 \times 2^{\rm L-16}$ is required. For example, $\mathrm{V21}_{24}$ corresponds to the following cell shifts in PANPHASIA:

$$\left.\begin{aligned}\Delta N_x &= 4 \times 12 \times 2^{21-16} = 1536,\\ \Delta N_y &= 2 \times 12 \times 2^{21-16} = 768,\\ \Delta N_z &= 0.\end{aligned}\right\} L = 21$$

$$\left.\begin{aligned}\Delta N_x &= 4 \times 12 \times 2^{22-16} = 3072,\\ \Delta N_y &= 2 \times 12 \times 2^{22-16} = 1536,\\ \Delta N_z &= 0.\end{aligned}\right\} L = 22$$

$$\left.\begin{aligned}\Delta N_x &= 4 \times 12 \times 2^{23-16} = 6144.\\ \Delta N_y &= 2 \times 12 \times 2^{23-16} = 1536,\\ \Delta N_z &= 0.\end{aligned}\right\} L = 23$$

This paper has been typeset from a TEX/LATEX file prepared by the author.