Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **515,** 1854–1870 (2022)
Advance Access publication 2022 July 22

https://doi.org/10.1093/mnras/stac1947

The DESI *N*-body simulation project – I. Testing the robustness of simulations for the DESI dark time survey

Cameron Grove [1★] Chia-Hsun Chuang [2,3★] Ningombam Chandrachani Devi,[4,5] Lehman Garrison [6] Benjamin L'Huillier [7] Yu Feng,[8] John Helly,[1] César Hernández-Aguayo [9,10] Shadab Alam,[11] Hanyu Zhang [12] Yu Yu [13] Shaun Cole,[1] Daniel Eisenstein,[14] Peder Norberg,[1,15] Risa Wechsler,[2] David Brooks,[16] Kyle Dawson,[3] Martin Landriau,[17] Aaron Meisner,[18] Claire Poppett,[19] Gregory Tarlé[20] and Octavio Valenzuela[4]

Affiliations are listed at the end of the paper

Accepted 2022 July 5. Received 2022 June 13; in original form 2021 December 2

ABSTRACT

Analysis of large galaxy surveys requires confidence in the robustness of numerical simulation methods. The simulations are used to construct mock galaxy catalogues to validate data analysis pipelines and identify potential systematics. We compare three *N*-body simulation codes, ABACUS, GADGET-2, and SWIFT, to investigate the regimes in which their results agree. We run *N*-body simulations at three different mass resolutions, 6.25×10^8, 2.11×10^9, and $5.00 \times 10^9\ h^{-1}\,\mathrm{M}_\odot$, matching phases to reduce the noise within the comparisons. We find systematic errors in the halo clustering between different codes are smaller than the Dark Energy Spectroscopic Instrument (DESI) statistical error for $s > 20\ h^{-1}$ Mpc in the correlation function in redshift space. Through the resolution comparison we find that simulations run with a mass resolution of $2.1 \times 10^9\ h^{-1}\,\mathrm{M}_\odot$ are sufficiently converged for systematic effects in the halo clustering to be smaller than the DESI statistical error at scales larger than $20\ h^{-1}$ Mpc. These findings show that the simulations are robust for extracting cosmological information from large scales which is the key goal of the DESI survey. Comparing matter power spectra, we find the codes agree to within 1 per cent for $k \leq 10\ h\,\mathrm{Mpc}^{-1}$. We also run a comparison of three initial condition generation codes and find good agreement. In addition, we include a quasi-*N*-body code, FastPM, since we plan use it for certain DESI analyses. The impact of the halo definition and galaxy–halo relation will be presented in a follow-up study.

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

1 INTRODUCTION

For many years *N*-body simulations have been used as a tool to explore the non-linear evolution of the distribution of matter in the Universe (Davis et al. 1985). Their use has been invaluable in creating mock galaxy catalogues to validate the results of surveys such as 2dFGRS, SDSS, DES, KiDS, and more (Cole et al. 1998; Jong et al. 2012; Rodríguez-Torres et al. 2016; DeRose et al. 2022). The Dark Energy Spectroscopic Instrument (DESI) will be performing surveys of unprecedented size over the next 5 yr, measuring tens of millions of galaxy spectra (DESI 2016). The large size of the DESI survey means that the statistical errors on key measured quantities will be small and therefore keen attention must be paid to systematic errors in all stages of the data collection and analysis. Measuring the size of systematic errors introduced by *N*-body simulations is imperative to understand their impact on mock galaxy catalogues and analysis pipelines.

One way to measure systematic errors for an individual *N*-body code is using convergence testing. The code can be run with progressively more accurate parameter choices until measured statistics no longer change (Power et al. 2003; DeRose et al. 2019). However, convergence testing is limited because it only provides information on the systematic errors which can be reduced by running a code with greater numerical accuracy. Different implementations or assumptions adopted by different codes could introduce systematic errors as well. Comparing the results from several different codes can indicate the level of control over systematic errors in converged runs of *N*-body simulation codes.

Code comparison projects measure the precision of *N*-body simulation codes when run from identical initial conditions (ICs). Measurements of the power spectrum, clustering, and halo properties can be compared to indicate the precision to which *N*-body simulations can estimate these quantities (Winther et al. 2015; Schneider et al. 2016; Garrison, Eisenstein & Pinto 2019).

This paper contains the results from multiple code comparison projects using several modern *N*-body simulation methods. Firstly, we compare the ICs created using several different codes, measuring the power spectrum along with particle velocity statistics. The main *N*-body code comparison project compares simulations run with

★ E-mail: cameron.grove@durham.ac.uk (CG); albert.chuang@utah.edu (CHC)

Table 1. The number of particles and particle masses of the simulation boxes. All are cubic boxes with a side length of 500 h^{-1} Mpc.

Simulation	Particle number	Particle mass/h^{-1} $M_\odot$
Low resolution	1296^3	5.00×10^9
Medium resolution	1728^3	2.11×10^9
High resolution	2592^3	6.25×10^8

ABACUS, FASTPM, GADGET-2, and SWIFT from identical ICs. The matter power spectrum is measured, along with dark matter halo properties, abundances, clustering, and power spectra. The precision of these results between different *N*-body codes and simulation resolutions is measured.

In Section 2, the ICs comparison is performed between three different codes. In Section 3, we discuss the different codes used to run the *N*-body simulation and we compare the results of the *N*-body simulations in terms of the matter power spectrum. In Section 4, we compare the halo catalogues produced from the different simulation codes, including measurements of the clustering, halo mass functions (HMFs), and matched halo properties. Finally, we summarize and conclude in Section 5.

2 INITIAL CONDITIONS CODES

The starting points for *N*-body simulations are the ICs. These define the positions, velocities, and masses of particles which are to be fed into the *N*-body simulation code. Analysis of the cosmic microwave background radiation suggests that the initial matter density field produced during inflation is a Gaussian random field (Planck Collaboration VI 2020). The properties of this field are determined by its power spectrum, which in turn is determined by the cosmology of the Universe. Different realizations of the same power spectrum are possible as the phases of the field are randomly generated (Abrahamsen 1997).

In this section, we discuss three different IC generators. The considered IC codes include those used in conjunction with the ABACUS (Garrison et al. 2021b) and FASTPM (Feng et al. 2016) *N*-body codes, in addition to a combination of two codes, PANPHASIA[1] (Jenkins 2013) and IC_GEN which we refer to as PANPHASIA in this text, see Section 2.1.2 for details.

The cosmology is set according to the flat Lambda cold dark matter Planck 2018 results (Planck Collaboration VI 2020) with the following parameters: the Hubble expansion rate at present time, $H_0 = 67.36$ km s^{-1}, $\Omega_c h^2 = 0.1200$, $\Omega_b h^2 = 0.0223$, $\Omega_\Lambda = 0.6834$, RMS linear density fluctuation, $\sigma_8 = 0.8$, and scalar spectral index $n_s = 0.9649$.

Considering the requirements of the DESI project, we run the simulations in boxes of side length, $L_{box} = 500\ h^{-1}$ Mpc with three different dark matter particle numbers: $N_p = 1296^3$, 1728^3, and 2592^3. The ICs are generated at redshift, $z_{ini} = 199$ with the input theoretical matter power spectrum generated using CAMB.[2] The details of the methods employed by each code are described below. We study the consistency between the codes in both the dark matter position and velocity distributions. The convergence test in each individual code is also performed using the three sets of resolutions. In Table 1, we provide the specifications of the simulations.

2.1 Code details

This section contains descriptions of the codes used to create ICs used in the comparisons in Sections 2.2.1 and 2.2.2.

2.1.1 ABACUS IC

The ABACUS IC code, called ZELDOVICH-PLT,[3] generates Zel'dovich approximation ICs, optionally applying particle linear theory (PLT) corrections (Marcos et al. 2006; Garrison et al. 2016) for the breaking of linear theory that occurs on small scales due to the discrete representation of the density field by particles. The PLT corrections are not enabled for purposes of this code comparison. The code is written in C++ and is designed to produce ICs in memory-limited environments by buffering the state to disc. The code uses double precision internally, and the positions are output as single-precision displacements on a lattice. The random number generator for the initial Gaussian density field outputs 64 bits and allows synchronizing the white noise at different resolutions; this capability is employed for the three resolutions of this IC code comparison. The input power spectrum normalization is recomputed by ZELDOVICH-PLT then scaled back to the requested redshift using the growth factor ratio from ABACUS's cosmology module. Modes are filled out to the Nyquist sphere. This is the ICs generator used by the ABACUS code, including the AbacusSummit project[4] (Maksimova et al. 2021).

2.1.2 IC_GEN/PANPHASIA

There are two components to this method of creating the ICs which has been labelled PANPHASIA. Firstly PANPHASIA describes a method of setting the phases of cosmological ICs. PANPHASIA describes a pseudo random number sequence mapped to a cosmological volume from which subsections can be taken to provide the phases for ICs (Jenkins 2013). This allows simple creation of resimulation ICs and also for the phases used in an *N*-body simulation to be shared so that others can recreate the same cosmological volume at any resolution without having to share the full phase information.

Once the phases have been generated using PANPHASIA, the ICs are generated using a proprietary code called IC_GEN. This uses second-order Lagrangian perturbation theory (2LPT, Jenkins 2010) to create the ICs particle distribution according to the phases specified by the location within PANPHASIA.

2.1.3 FastPM IC

While not originally designed as an IC code, FastPM does compute an initial particle state with the traditional 2LPT-on-a-mesh approach. In addition, the IC code in FastPM supports producing constrained Gaussian realizations.

Particles are first placed on a uniform grid with N_c particles per side, then the first-order LPT and 2LPT displacement ($\mathbf{s}_{1,2}$) and velocity terms ($\mathbf{v}1, 2$) are calculated from a potential induced from a Gaussian white noise field $g(\mathbf{x})$ according to an isotropic linear power spectrum $P_{lin}(k)$. The white noise field is sampled on a finite mesh of $B_{LPT}N_c$ samples per side from a pseudo-random generator compatible to the one used by Gadget's N-GenIC (Springel et al. 2005; Angulo et al. 2012).

[1] http://icc.dur.ac.uk/Panphasia.php
[2] https://camb.info/
[3] https://github.com/abacusorg/zeldovich-PLT
[4] https://abacussummit.readthedocs.io

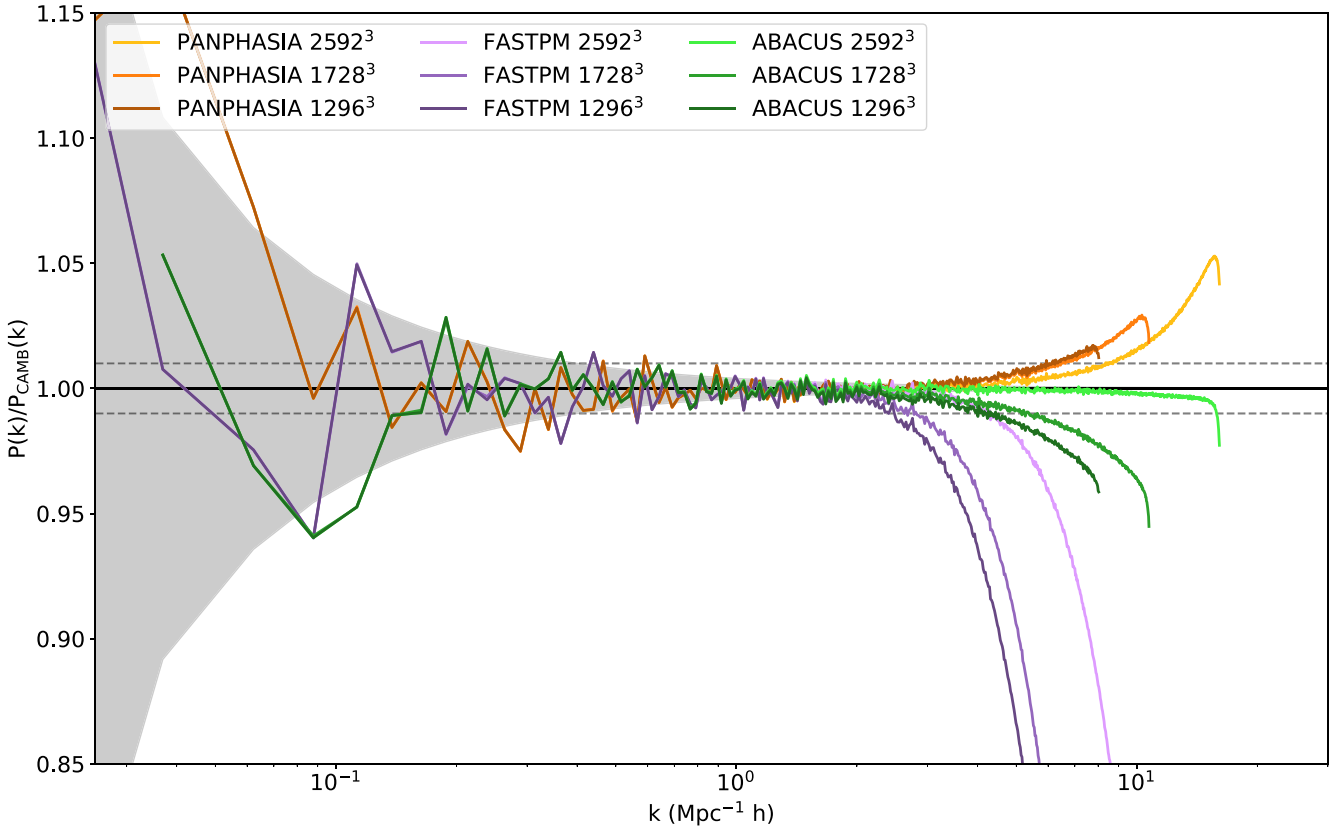

Figure 1. A comparison of the power spectrum of ICs created at redshift $z = 199$, using different codes and resolutions. The power spectra are plotted as a ratio to the input theoretical power spectrum from CAMB. At high k differences emerge between the input and the measured power spectra. At low k, the sample variance causes noise in the measured power spectrum. The grey envelope shows the theory 1σ sample variance. The dashed lines indicate ± 1 per cent.

The choice of $B_{\rm LPT}$, the finite differentiation operators, and the finite Laplace operator all affect the initial particle state, even though the effect on the late-time non-linear field can be quite minimal.

2.2 Comparison of ICs

In this section, we compare the ICs in terms of particle and velocity distributions.

2.2.1 *Comparison of particle distribution*

The power spectrum is a common measure for analysing both ICs and evolved N-body simulations. It measures the strength of the matter density contrast at different scales.

The power spectrum, $P(k)$ is defined as the Fourier transform of the autocorrelation function, $\xi(r)$ which in turn is derived from the matter overdensity, $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1$,

$$P(k) = \int d^3\mathbf{r}\, e^{-i\mathbf{k}\cdot\mathbf{r}}\xi(r), \tag{1}$$

$$\xi(r) = \langle\delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r})\rangle = \frac{1}{V}\int d^3\mathbf{x}\,\delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r}). \tag{2}$$

ICs sample the matter density field producing a set of particles. This limited number of particles means that there is a maximum k above which the power spectrum cannot be represented in the ICs. This is the Nyquist frequency. For a box with sidelength L and N^3 particles the wavenumber associated with the Nyquist frequency is

$$k_{\rm Nyquist} = \frac{\pi N}{L}. \tag{3}$$

The Nyquist wavenumber, and therefore the Nyquist frequency, increases with number of particles allowing us to compare to the reference power spectrum at higher k.

The power spectra are measured using NBODYKIT[5] (Hand et al. 2018) using a triangular shaped cloud mass assignment scheme and grid sizes of 2592^3, 3456^3, and 5184^3 were used for the low, medium, and high resolution ICs, respectively. Differences near the Nyquist frequency are observed when measuring the power spectra with different grid sizes and aliasing is observed when using a grid size which does not evenly divide the particle distribution. Using large grid sizes which are multiples of the particle numbers mitigates these effects. The power spectra are truncated at the Nyquist frequency for each resolution.

Fig. 1 shows the power spectra measured for all ICs codes and also for different resolutions, compared to the reference power spectrum. The reference power spectrum is the theoretical linear power spectrum from CAMB at redshift, $z = 199$. The measured power spectra differ from the reference power spectrum at both large (low k) and small scales (high k). The large-scale differences are mainly due to sample variance and the limited number of large-scale

[5] https://nbodykit.readthedocs.io/en/latest/

modes in a particular realization may lead to disagreement in the power spectrum measurement from the reference P_k at these scales. The grey envelope shows the theory sample variance. This variance decreases as k^{-1} which is consistent with the theory expectation using Poisson noise where the number of modes is proportional to k^2.

There is only one distinct line for each code at low k, this is because each code uses a different phase realization which is consistent between different resolutions. At small scales, close to the Nyquist frequency, the systematic resolution effects are different between codes. The ICs created using PANPHASIA do not disagree with the reference power spectrum by more than 5 per cent in all resolutions up to the Nyquist frequency. The FASTPM ICs have much lower power than the reference power spectrum for $k > k_{\rm Nyquist}/2$. The ABACUS ICs have the best agreement with the reference power spectrum.

It should be noted that it is not imperative for ICs to exactly fit the reference power spectrum at the smallest scales in order to create a realistic N-body simulation. The majority of the power at low redshift at small scales is due to collapse of larger modes and therefore small differences in the power spectrum around $k_{\rm Nyquist}$ are mostly unimportant to the growth of structure at these scales (Neyrinck & Yang 2013).

2.2.2 *Comparison of velocity statistics*

The power spectrum is only a measure of the displacements of the particles produced by a IC code. The particle velocities are just as important as the displacements and affect the rate of growth of structure as the simulation progresses. In methods of generating the ICs, such as the Zel'dovich approximation (Zel'Dovich 1970) and 2LPT (Crocce, Pueblas & Scoccimarro 2006), the particle velocities are not free to vary but are fixed by the particle displacements. For example in the Zel'dovich approximation a particle's velocity is in the same direction as, and proportional to, the displacement from the particle's gridpoint. For this comparison project, we thought it prudent to check the consistency between particle velocities from different IC codes. We investigated the pairwise velocity dispersion between the codes and also compared to 10 realizations of PANPHASIA ICs. There was no theoretical estimate generated therefore we aimed to verify that there were similar results between the codes with the differences being consistent with the sample variation from the 10 PANPHASIA realizations.

The pairwise velocity dispersions are calculated by taking the pairwise velocity between pairs of particles separated by a distance r along with the component of the velocity parallel and perpendicular to the vector between the particles.

For particles with positions and velocities $(\mathbf{x_1}, \mathbf{v_1})$ and $(\mathbf{x_2}, \mathbf{v_2})$ the parallel and perpendicular components are

$$v_\parallel = \frac{(\mathbf{x_1} - \mathbf{x_2}) \cdot (\mathbf{v_1} - \mathbf{v_2})}{|\mathbf{x_1} - \mathbf{x_2}|} \tag{4}$$

and

$$v_\perp = \sqrt{|\mathbf{v_1} - \mathbf{v_2}|^2 - v_\parallel^2} \tag{5}$$

respectively. This gives velocity dispersions in the parallel and perpendicular directions of

$$\sigma_\parallel^2(r) = \left\langle v_\parallel^2(r) \right\rangle - \left\langle v_\parallel(r) \right\rangle^2 \tag{6}$$

and

$$\sigma_\perp^2(r) = \left\langle v_\perp^2(r) \right\rangle - \left\langle v_\perp(r) \right\rangle^2. \tag{7}$$

The isotropy of the universe implies that the second term of equation (7) is equal to zero.

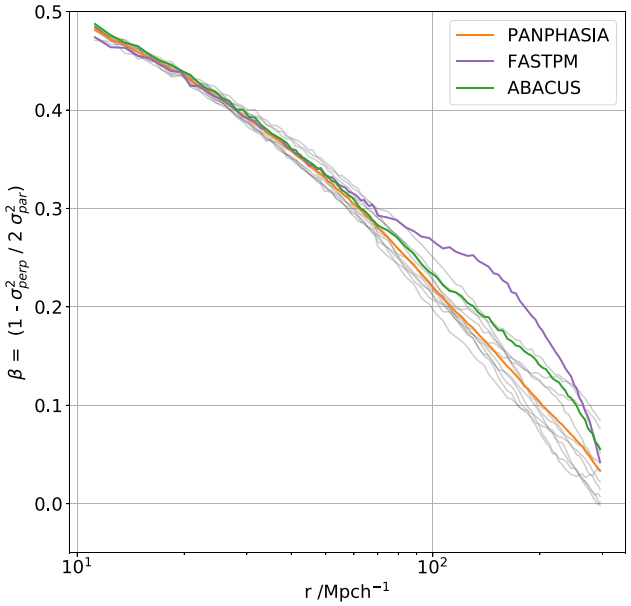

Figure 2. A plot showing the velocity anisotropy as a function of interparticle distance. The coloured lines show the different codes meanwhile the grey lines show the results for 10 PANPHASIA realizations to investigate sample variance. The codes agree well at small distances while at large r the differences are dominated by sample variance. The only outlier is FastPM which has a higher anisotropy than any of the other realizations around 150 h^{-1} Mpc.

The velocity anisotropy parameter

$$\beta = 1 - \frac{\sigma_\perp^2}{2\sigma_\parallel^2} \tag{8}$$

measures the relative size of the parallel and perpendicular velocity dispersions.

Fig. 2 shows that the velocity anisotropy, as a function of pairwise interparticle distance, is mostly consistent between the codes and at large scales the differences are mostly explained by sample variance. There is excellent agreement between the codes at small interparticle distance. FastPM is the only outlier, having a slightly higher anisotropy than the other realizations at around $r = 150\ h^{-1}$ Mpc.

It was decided to use PANPHASIA ICs for the N-body code comparison project in the next section. This was because PANPHASIA had reasonable agreement with the reference power spectrum close to the Nyquist frequency in all resolutions and PANPHASIA has the ability to easily share phase information to be used by other ICs codes in the future. These tests do not show whether the IC generators are fully accurate in absolute terms, and the impact of IC differences on evolved simulations is not measured. However these results show that the PANPHASIA ICs are a reasonable choice to use for the code comparison project when compared to the other options. We have publicly released the IC files so that further comparisons can be made if there is interest, this is detailed in the Data Availability section.

3 *N*-BODY CODES

This section details the N-body code comparison project. We run ABACUS, SWIFT, GADGET-2, and FASTPM from $z = 199$ to $z = 0$ at the three resolutions shown in Table 1. Comparisons between the codes are made in the matter power spectrum and in the

dark matter halo properties, along with comparisons of the halo clustering.

The ICs for this code comparison project were created using PANPHASIA and IC_GEN from the previous section; all codes used identical initial particles. The cosmology used is described at the start of Section 2 and the resolutions of the simulation boxes are shown in Table 1.[6]

Codes were run by separate groups and the code parameters were decided upon by those groups.

3.1 Code details

3.1.1 ABACUS

ABACUS is a high-force accuracy N-body code that solves the far-field force with a high-order multipole method and the near field with direct summation on GPUs (Garrison et al. 2021b). Typical RMS force accuracy is 10^{-5}–10^{-6}. All particles in ABACUS share a single, global time-step, which is chosen at the beginning of each time-step. ABACUS was recently used to run the AbacusSummit simulations (Maksimova et al. 2021), and the ABACUS realization of the Euclid code comparison simulation was presented in Garrison et al. (2019).

The numerical parameters used in this code comparison are the same as those used in AbacusSummit. Specifically, spline force softening, fixed in proper coordinates, was used such that the $z = 0$ value was 1/40th of the interparticle spacing (with the early-time softening capped to 0.3 in the same units). This choice was tested using scale-free simulations in Garrison, Joyce & Eisenstein (2021a). The multipole order was $p = 8$, with 405^3 cells, and the time-step parameter was set to $\eta_{\rm Accel} = 0.25$. This is mildly more conservative than ABACUS's run of the Euclid code comparison simulation, which used $\eta_{\rm Accel} = 0.3$. Garrison et al. (2021b) show that a time-step value of $\eta = 0.25$ produces subpercent shifts in the two-point clustering for all scales larger than twice the softening length.

3.1.2 GADGET-2

GADGET-2 (Springel 2005) combines a tree and a particle mesh for its gravity scheme. The first-order moments are used to calculate the tree forces and a fixed opening angle criterion is used to decide when to divide up the tree cells. The particle positions and velocities are updated in a leapfrog scheme and the global time-steps are evenly spaced in $\log(a)$ with dynamic shorter time-steps for particles undergoing large accelerations. A mesh size of N^3 was used for the simulation with N^3 particles. The comoving Plummer equivalent softening length used was 1/25th of the interparticle spacing. The parameters used to define the time-stepping and force accuracy were MaxRMSDisplacementFac = 0.2, MaxSizeTimestep = 0.01, ErrTolTheta = 0.5, and ErrTolForceAcc = 0.0025.

3.1.3 SWIFT

SWIFT[7] (Schaller et al. 2016) is a hydrodynamics and gravity code which uses a task based parallelization strategy. It is primarily designed for high-resolution SPH simulations of galaxy formation; however, SWIFT can also be used for dark matter only simulations of large-scale structure. The gravity scheme in SWIFT has three levels. At large scales a Fourier mesh is used to calculate the force on particles and account for the periodicity of the simulation box; at small scales the fast multipole method (FMM, Dehnen 2014) is used to calculate forces between particles in octree cells; at the smallest scales direct summation is used. A fixed opening angle criterion is used to calculate the size of the octree cells used in the FMM. A time-dependent force softening is applied to the particles, the forces are softened according to a spline kernel. The global time-steps are evenly spaced in $\log(a)$ and adaptive time-stepping is used for particles with large accelerations in the same way as in GADGET-2 (Theuns et al. 2015).[8] A Fourier mesh size of 648^3 was used in the 1296^3 particle and 2592^3 particle simulations and 864^3 was used in the 1728^3 particle simulation. These values were chosen as they were the largest allowed by the code which would not introduce aliasing with the initial particle grid. The Plummer equivalent comoving softening length was 1/25th of the interparticle spacing.

3.1.4 FASTPM

FASTPM[9] (Feng et al. 2016) is an approximated particle mesh N-body solver. By modifying the kick and drift factors in time integration, it enforces correct linear displacement evolution on large scales, and meanwhile reduces the halo stochasticity when the number of time-steps is small (Feng et al. 2016). The cost is extremely low thanks to the small number of time-steps required for satisfying statistics. The accuracy can be mainly controlled by two options, B, the mesh size over particle number per dimension in the force calculation, and the time-steps usually specified by the number of steps T and the spacing scheme. In this comparison project, we use $B = 2$ and $T = 46$, with linear spacing in the scale factor. FastPM simulations are useful for estimating covariance matrices and other applications for which the accuracy is not a critical factor. A summary of the FastPM simulations prepared by the DESI Cosmological Simulations working group will be presented in Ding et al. (in preparation).

3.2 Comparison of particle distribution

We first begin by examining the power spectrum dependence on the mass resolution for each code separately. As seen in Fig 3, this dependence is similar for all the codes tested. There is very little change in the power spectra at large scales and at small scales there are differences of 1–2 per cent. The higher the mass resolution, the higher the power at small scales (except with FASTPM).

We then proceed to compare the power spectrum results among codes. Fig. 4 shows all of the power spectra relative to the ABACUS runs. There is broadly good agreement between all of the codes except FASTPM with differences below 1 per cent at $k = 2\,h\,{\rm Mpc}^{-1}$. This agrees with the results found in other code comparison projects (Schneider et al. 2016; Garrison et al. 2019). We do not see the discrepancy between GADGET-2 and linear theory at large scales as seen in Schneider et al. (2016), although this work employs GADGET-2 and Schneider et al. employed GADGET-3. Apart from FASTPM, at $k = 10\,h\,{\rm Mpc}^{-1}$ the fractional differences between the power spectra are below 1 per cent.

All codes agree on the low-k power spectrum amplitude to within 0.1 per cent, but to ascertain that this value is that predicted by

[6]The Panphasia descriptor describing the simulation volume is [Panph1,L21,(1133107,236124,673886),S81,CH1311607586,DESI_IC_v1].
[7] http://swift.dur.ac.uk/
[8]https://gitlab.cosma.dur.ac.uk/swift/swiftsim
[9]https://github.com/rainwoodman/fastpm

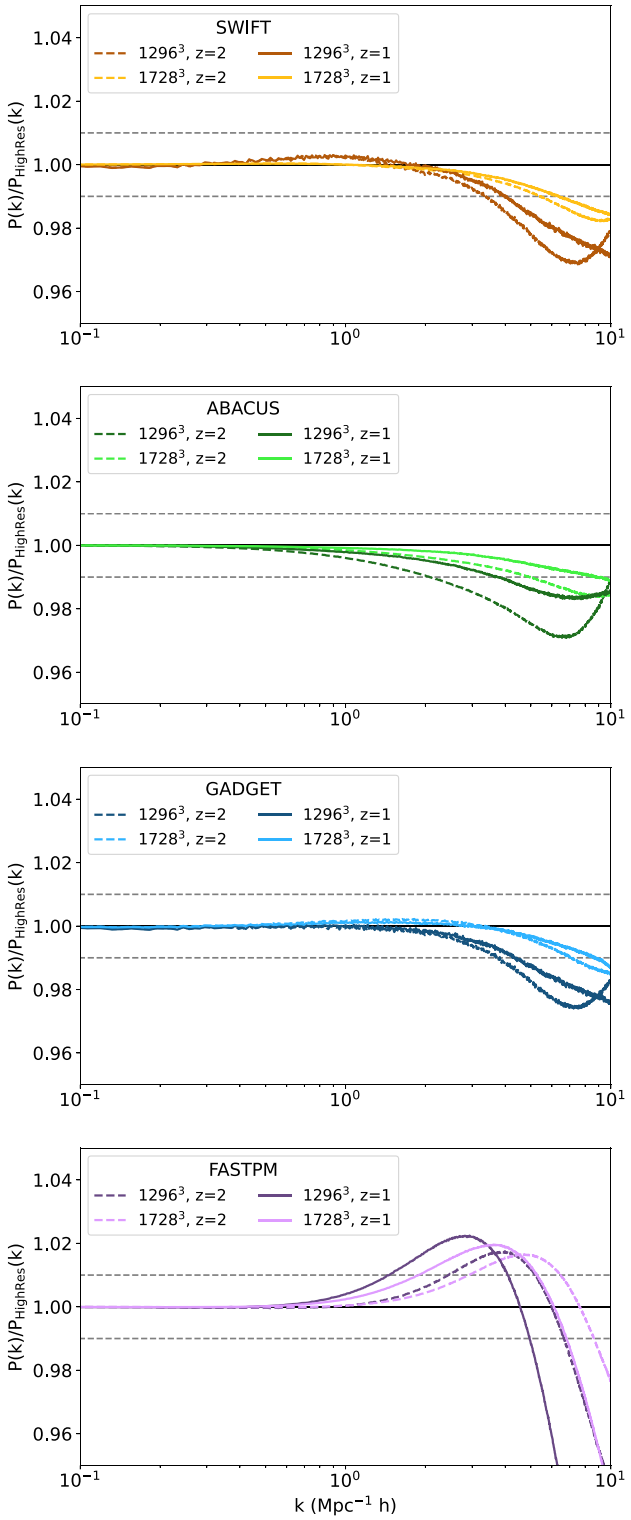

Figure 3. The matter power spectrum of the low- and medium-resolution runs relative to that of the high-resolution run for each code at $z = 1$ and 2. SWIFT, ABACUS, & GADGET-2 show similar behaviour where the power spectrum decreases by up to 2 per cent at $k = 10\ h\,\mathrm{Mpc}^{-1}$ when decreasing the resolution from 2592^3 to 1728^3 particles or from 1728^3 to 1296^3 particles. The lower resolution FASTPM power spectra show an increase relative to the high-resolution case at $k = 3\ h\,\mathrm{Mpc}^{-1}$ before decreasing sharply above $k = 5\ h\,\mathrm{Mpc}^{-1}$.

perturbation theory, the 2LPT estimates of the $z = 1$ and $z = 2$ power spectra were measured by using IC_GEN, the code used to create the ICs for the simulations. Using a low-output redshift for the same phases gives us the estimate of the matter distribution at low redshift from which the power spectrum can be measured. The results of this test are shown in Fig. 5. We also attempted to compare to linear theory by rescaling the power spectrum of the ICs by the linear growth factor; however, this did not reproduce the large-scale power spectrum to subpercent accuracy, as shown in the dashed lines in Fig. 5. The reason for this difference is likely to be mode coupling. In addition, using the Zel'dovich approximation instead of 2LPT produced a curve which was less smooth and therefore we used 2LPT for the low-redshift perturbation theory power spectrum. Perturbation theory is only accurate for small-density perturbations, which corresponds to very large scales at low redshift. We have found that the simulation power spectra and the 2LPT power spectra asymptotically converge at low k for all the codes. Furthermore, this exercise helped identify a previously unknown error in the large-scale growth in the SWIFT code; a difference of around 0.5 per cent in the large-scale power spectrum which was caused by incorrect force integration. The error was corrected, and all results in this work use the modified code.

4 DARK MATTER HALO COMPARISON

Much of the value which comes from running large N-body simulations does not come from the complete distribution of dark matter particles, but instead comes from grouping the particles into overdensed regions called dark matter haloes because these are the sites that host galaxies (Benson et al. 2000; Peacock & Smith 2000; Conroy & Wechsler 2009). Using and storing full particle data can become prohibitively expensive for large N-body simulations therefore halo catalogues are often the most useful data product to be produced from an N-body simulation. Observations and simulations suggest galaxies reside within dark matter haloes and therefore they are essential to connect N-body simulations to observable results (Wechsler & Tinker 2018).

Halo catalogues can be used to create mock galaxy catalogues which provide useful information on the expected observational results for different cosmologies, along with providing mock data which is essential for testing analysis pipelines, e.g. see Cole et al. (1998), Rodríguez-Torres et al. (2016), and DeRose et al. (2022).

In this section, we perform the comparison of halo properties, mass functions, and clustering. Haloes are identified with the help of the phase-space temporal halo finder ROCKSTAR with force resolution parameters shown in Table 2 (Behroozi, Wechsler & Wu 2013),[10] and default parameters otherwise, including `STRICT_SO_MASSES=0`. ROCKSTAR finds haloes by using a friends-of-friends algorithm in six phase-space dimensions (three spatial and three velocity dimensions). Groups and subgroups are found using an adaptive hierarchical method which progressively reduces the linking length between particles. Setting `STRICT_SO_MASSES=0` affects the mass definition of haloes. The halo mass in this case is defined only using particles within the friends-of-friends group, rather than also including particles which exist outside the group but inside the spherical overdensity.

We do not include FASTPM in this comparison of halo properties and clustering. This is because performing abundance matching is outside the scope of this paper, and the FASTPM results without

[10] https://bitbucket.org/gfcstanford/rockstar

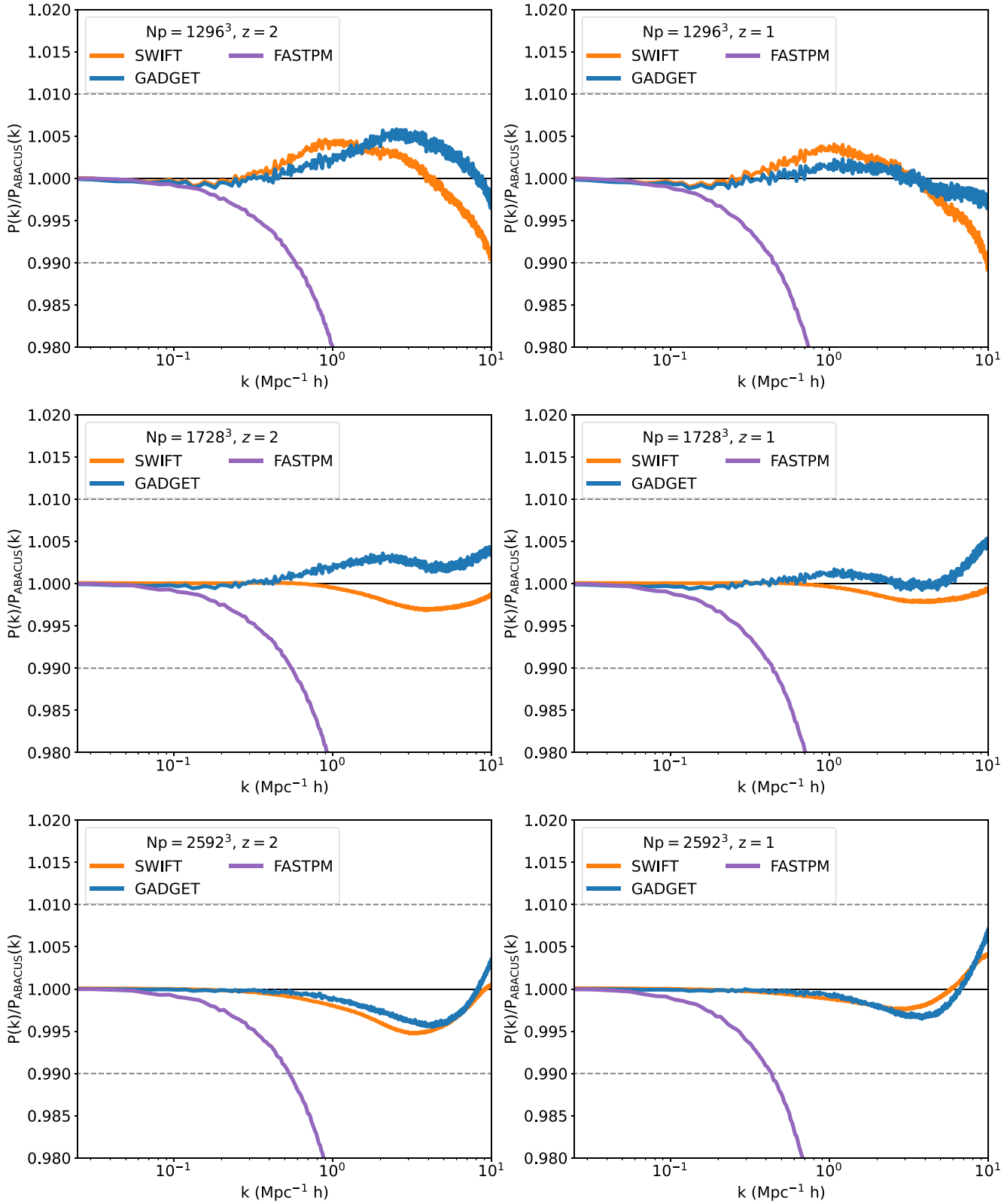

Figure 4. A comparison of the matter power spectrum between the different codes at $z = 1$ and $z = 2$ in all three resolutions. The power spectrum is plotted relative to ABACUS at each redshift and resolution. All the codes agree to within 0.1 per cent at $k < 0.1\ h\,\mathrm{Mpc}^{-1}$. Differences emerge at high k but, excepting FASTPM, these are within 1 per cent at $k = 10\ h\,\mathrm{Mpc}^{-1}$.

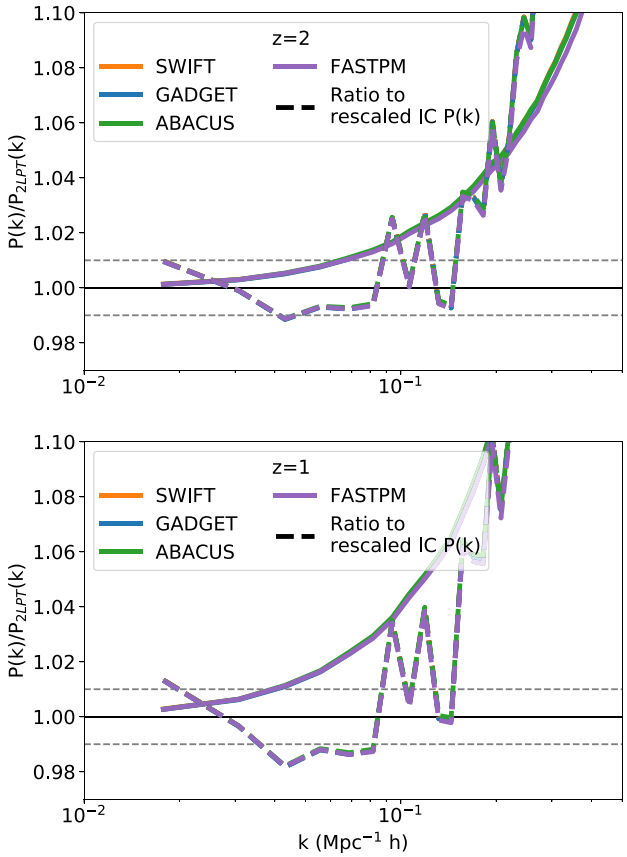

Figure 5. A comparison of the power spectra to perturbation theory at $z = 1$ and $z = 2$. The perturbation theory power spectrum was found using second-order perturbation theory (2LPT). The power spectra should asymptotically agree with 2LPT at low k. Coloured dashed lines show the ratio of the power spectra with ICs rescaled by the growth factor which do not agree as closely as 2LPT.

Table 2. ROCKSTAR force resolution. This table shows the force resolution of the simulations in comoving h^{-1} Mpc, as input to the ROCKSTAR halo finder. Haloes whose centres are closer than the force resolution are considered to be unresolved. The minimum number of particles considered to be a halo seed for 1296^3, 1728^3, and 2592^3 resolutions are 5, 10, and 10, respectively.

Simulation	ABACUS	GADGET-2	SWIFT
1296^3	0.015	0.015	0.015
1728^3	0.0072	0.011	0.011
2592^3	0.0048	0.0077	0.0077

abundance matching would be unrealistically discrepant compared with how the code is used in practice.

Our tests involving dark matter haloes are performed at $z = 1$. We do not observe significant redshift evolution in our code or resolution comparisons when measuring HMFs and halo clustering at $z = 0$ and $z = 2$.

4.1 Halo properties

We execute a detailed halo-to-halo comparison by matching haloes between N-body simulations and comparing their properties. There is no guarantee that haloes will be formed in identical places, but since all simulations share the same initial particles, in many cases a robust match can be identified. Still, matching haloes provides a challenge as there is no unique identifier for each halo which is well defined across several different simulation runs.

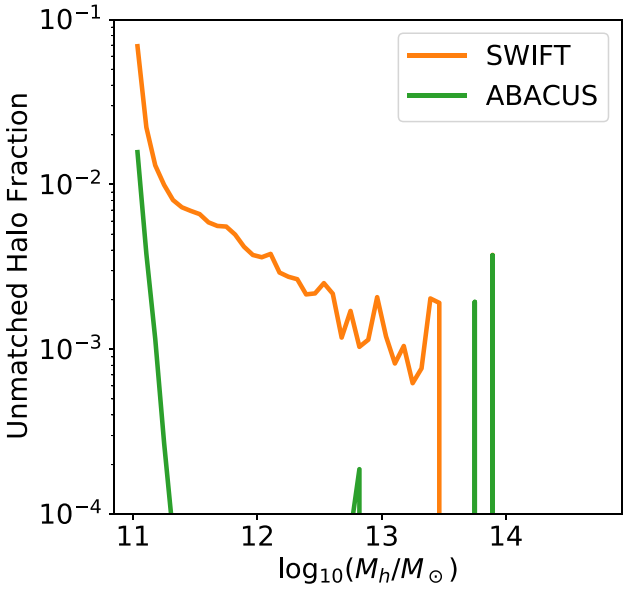

Figure 6. The fraction of haloes which remain unmatched as a function of halo mass in the ABACUS and SWIFT catalogues when matching haloes between the high-resolution catalogues. The majority of haloes are matched across simulations.

We use a nearest neighbour approach to match haloes between snapshots from simulations run with different codes and resolutions. Each halo in one catalogue was matched to the halo with the nearest position in the other catalogue. All haloes with a mass smaller than $5 \times 10^{10}\ h^{-1}\,\mathrm{M}_\odot$ at all resolutions were removed to avoid spurious matches and an upper distance limit of $0.25\ h^{-1}$ Mpc was applied for haloes to be considered as matched. The matched halo properties were binned by halo mass.

Fig. 6 shows the fraction of unmatched haloes in each catalogue in a comparison between the high-resolution SWIFT and ABACUS simulations. This fraction increases as halo mass decreases but is only greater than 5 per cent of haloes for the lowest mass SWIFT haloes at $M_h = 10^{11}\ h^{-1}\,\mathrm{M}_\odot$. This gives us confidence that there are not systematic effects being introduced by the nature of the halo matching algorithm.

The halo properties which are compared between matched catalogues are defined as follows. Halo separation is the distance between matched haloes in units of h^{-1} Mpc, as discussed above this cannot be greater than $0.25\ h^{-1}$ Mpc for any matched pair. M_{200b} is defined as the mass contained within a sphere around the halo such that the sphere has a density of 200 times the background density of the universe. $r_{\rm vir}$ is the radius of the halo which is similarly defined in terms of a spherical overdensity. In this case the overdensity is Δ_c times the critical density of the universe. Δ_c is the solution for a virialized cluster which is $18\pi^2$ for a critical universe but can vary with redshift otherwise. $M_{\rm vir}$ is the mass contained within $r_{\rm vir}$ (Peebles 1980; Eke, Cole & Frenk 1996; Bryan & Norman 1998). $r_{\rm s}$ is the scale radius of the halo, defined in terms of the density profile (Navarro, Frenk & White 1997). $\sigma_{\rm v}$ is the halo particle velocity dispersion in physical $\mathrm{km\,s^{-1}}$. $|v|$ is the speed of the halo in physical $\mathrm{km\,s^{-1}}$. Finally $\theta_{\rm v}$ is the difference in angle of the velocity vectors of

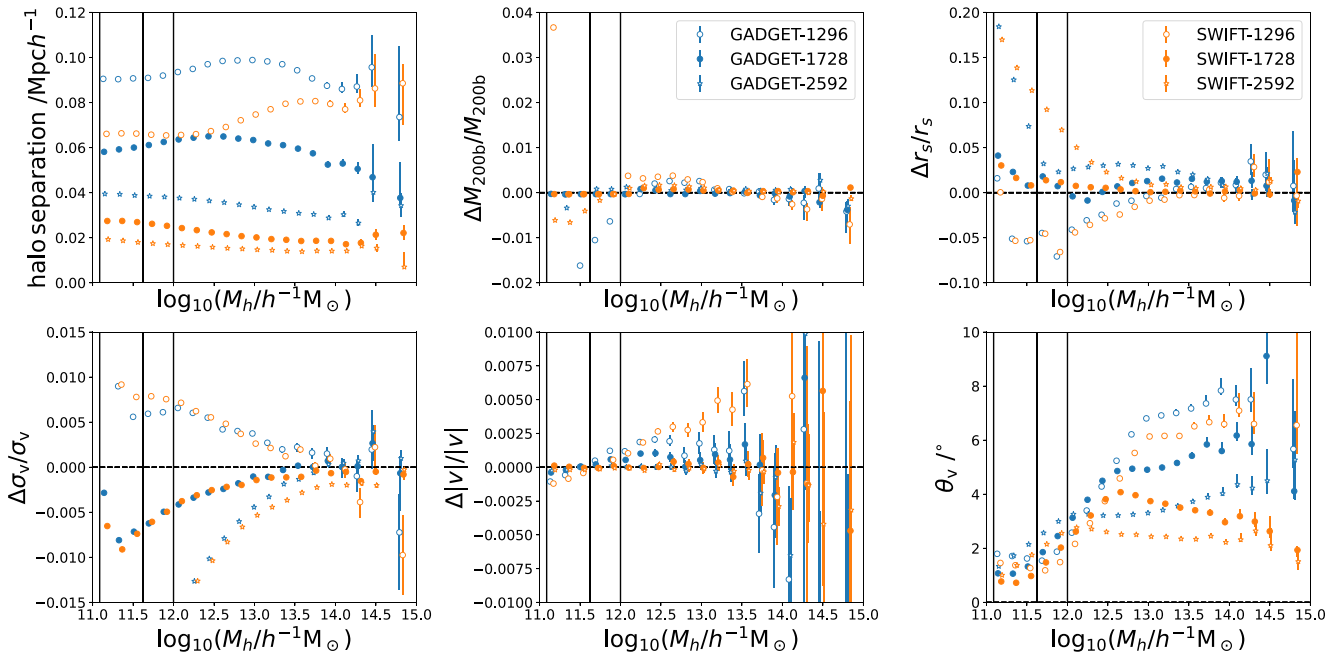

Figure 7. A comparison of the halo properties for haloes matched between ABACUS and the other codes at fixed resolution. On the top row from left to right the panels show: (i) The distance between matched halo centres in h^{-1} Mpc. (ii) The fractional difference in halo mass, as measured within a sphere of density 200 times the background density. The fractional difference in virial mass is identical to this panel, the fractional difference in virial radius is also identical to this panel but multiplied by a factor of three. (iii) The fractional difference in the scale radius of matched haloes. On the lower row, from left to right, the panels show: (i) The fractional difference in the velocity dispersion of matched haloes. (ii) The fractional difference in the magnitude of the velocities of the matched haloes. (iii) The difference in the direction of the velocities of matched haloes in degrees. Hollow circles, filled circles, and stars represent the comparison for the 1296^3, 1728^3, and 2592^3 simulations, respectively. The results are colour coded differently for each simulation code. GADGET-2: blue, SWIFT: orange, and ABACUS is used as the reference simulation. Vertical solid lines indicate the halo mass at which haloes contain 200 particles in the high-, medium-, and low-resolution simulations moving from left to right.

the matched haloes in degrees. With the exception of halo separation, $|v|$, and θ_v, which we compute ourselves, all of these properties are computed by ROCKSTAR.

Fig. 7 shows how the matched halo properties vary between ABACUS and the other codes at fixed resolution. The low, medium, and high resolution cases are shown as hollow circles, solid circles, and stars, respectively, along with error bars which show the error in the mean of the matched property within each mass bin.

In general, the matched properties show greater agreement at high masses and at high resolution. This is where the haloes contain a greater number of particles and therefore it is to be expected that these haloes are less susceptible to variation due to differences between the codes which are primarily in the small-scale forces. The halo separation is roughly constant across halo mass but varies for different codes and resolutions. The mean halo separation varies between 10 and 100 h^{-1} kpc; SWIFT, and ABACUS agree to within three softening lengths at high and medium resolution, while GADGET-2 has two to three times worse agreement.

SWIFT and GADGET show agreement with ABACUS to within 1 per cent in the halo mass and radius at medium and high resolution for all masses. At low resolution differences emerge for halo masses below 10^{12} h^{-1} $\mathrm{M}_\odot$.

r_s shows a similar pattern of better agreement at higher halo masses. The differences at masses below 10^{12} h^{-1} $\mathrm{M}_\odot$ are at the 10 per cent level and do not show clear trends with code or resolution, except that SWIFT and GADGET show progressively lower Δr_s at lower resolutions.

As resolution decreases, SWIFT and GADGET-2 have increasing $\Delta\sigma_v$ relative to ABACUS.

The halo velocities show larger differences in both magnitude and direction between the codes for high-mass haloes. There are large uncertainties on these measurements due to the limited number of high-mass haloes within each catalogue.

Fig. 8 shows how the matched halo parameters vary with simulation resolution when keeping the simulation code fixed.

There is a clearer trend in the halo separation with halo mass when comparing with each code fixed, the separation is lower for higher mass haloes and is also lower at higher resolution. SWIFT and GADGET-2 show larger halo separation at different resolutions compared to ABACUS.

The halo masses are systematically smaller by around 0.5 per cent for all codes and resolutions at halo masses above 10^{13} h^{-1} $\mathrm{M}_\odot$. The same feature is seen in the virial radii where the difference is 0.2 per cent. For lower mass haloes the opposite effect is seen, halo masses are larger for the lower resolution simulations and this difference grows when going to lower halo masses and lower resolutions. This indicates that at lower resolutions mass is systematically moved from high-mass haloes into low-mass haloes when compared to higher resolution simulations. The origin of this effect could be that the halo finder is more likely to split off haloes at low resolutions due to a lower number of particles with which to make friends-of-friends connections.

r_s is consistent between high-mass haloes at all resolutions for all the codes. Below 10^{12} h^{-1} $\mathrm{M}_\odot$, r_s becomes poorly matched as resolution is varied.

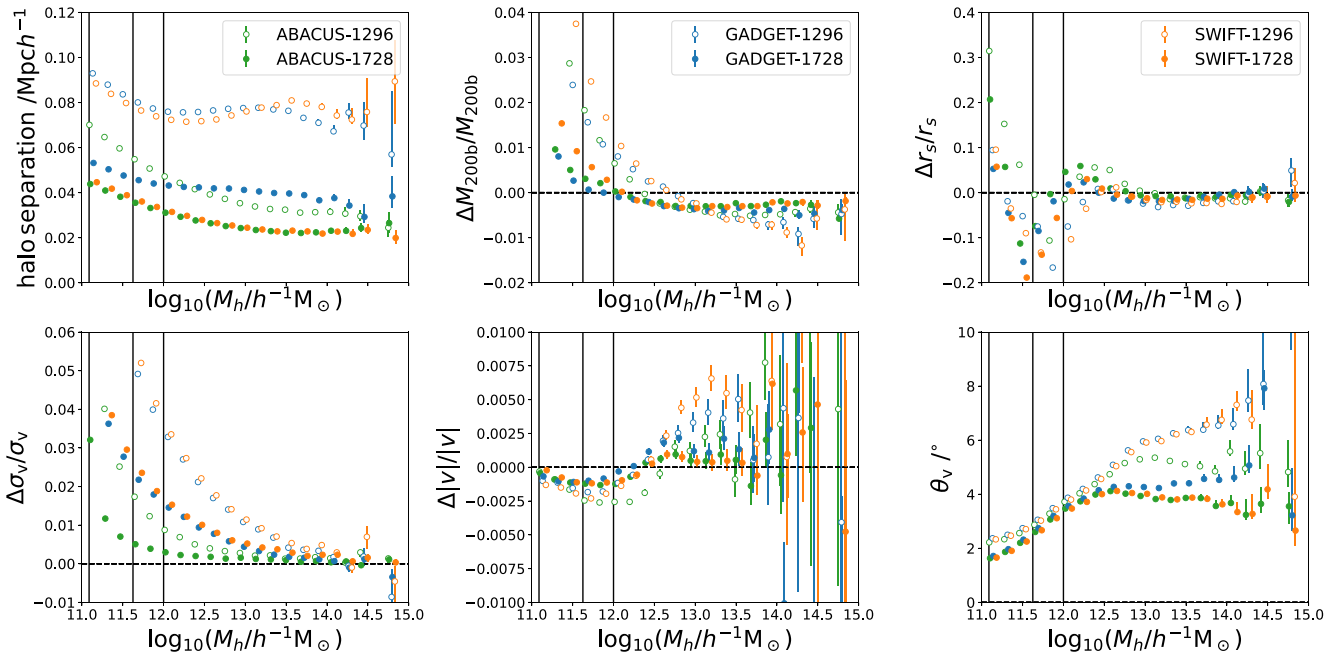

Figure 8. As for Fig. 7 but for haloes matched to the high-resolution simulation for each code. Note the change in range plotted in the upper-right and lower-left panels.

The velocity dispersion, σ_v, increases for lower mass haloes at lower resolutions. This effect can be seen for all the codes but is most strongly present in SWIFT, and GADGET-2.

In summary, the halo properties agree well between the different codes at fixed resolution, there is better agreement for haloes with mass greater than $10^{13}\ h^{-1}\,\mathrm{M}_\odot$ than lower mass haloes.

Lower simulation resolution systematically biases high-mass haloes to have slightly lower mass and radius than those matched from higher resolution simulations. The opposite effect is observed at low masses. In general high-mass haloes show greater agreement in the matched halo parameters than low-mass haloes.

4.2 Halo mass function

The HMF describes the number density of dark matter haloes in different mass intervals. Here, we have defined the HMF as $\mathrm{d}n/\mathrm{dlog}_{10}(M_{\rm vir})$ where n is the number density of haloes in units of $\mathrm{Mpc}^{-3}\,h^3$ and $M_{\rm vir}$ is in units of $h^{-1}\,\mathrm{M}_\odot$.

The HMF has a strong effect on the number density and clustering of galaxy catalogues created using the halo catalogue, particularly at fixed mass. It is therefore, important for the HMF to not change significantly between simulations run with different codes, as this could indicate underlying systematic errors.

In addition, we investigate the effect of simulation resolution on the HMF. Mass resolution has the greatest effect at the low-mass end of the HMF where haloes contain fewer particles. It is important to find the cutoff mass below which the HMF shows significant differences due to simulation resolution.

Fig. 9 shows the HMFs measured from the different codes at fixed resolution at $z = 1$, relative to the ABACUS HMF. The error bars represent the estimated standard deviation in the HMF difference between phase matched simulations which are calculated using a jackknife method, see equation (14) in Section 4.3 for more details. Differences between the codes are greater at high and low masses. The HMFs agree at the 1 per cent level between 10^{12} and $10^{13.5}\ h^{-1}\,\mathrm{M}_\odot$ for all resolutions. The HMFs do not differ by greater than 10 per cent at masses up to $10^{14.5}\ h^{-1}\,\mathrm{M}_\odot$, these differences are consistent with noise and not biased in a particular direction. At low masses there is different behaviour between the codes at each resolution. There is better agreement on the low-mass end of the HMF at low resolution than high resolution.

Fig. 10 shows the change in HMF with resolution for the ABACUS code at $z = 1$. The effect of changing resolution is larger than changing code. The halo catalogues from different resolution simulations have similar HMFs for high-mass haloes. The medium-resolution simulation HMF differs from the high-resolution simulation HMF by more than 1 per cent below $10^{12}\ h^{-1}\,\mathrm{M}_\odot$. The low-resolution HMF differs by a greater amount than the medium resolution HMF, with greater than 1 per cent disagreement below $10^{12.5}\ h^{-1}\,\mathrm{M}_\odot$. Testing the same comparison with a halo catalogue produced from a subsampled version of the high-resolution simulation produces a similar relationship which indicates that discreteness effects in the halo finder are not responsible for this systematic difference.

4.3 Halo clustering: correlation function

The clustering of dark matter haloes is one of the most important statistics to be produced in an N-body simulation. Galaxy clustering is closely tied to halo clustering, especially at large scales, therefore ensuring accurate and robust halo clustering estimates is essential in order to have confidence in mock data produced from simulations.

In this section, we compare the halo clustering from simulations run with different codes and at different resolutions. We compare the clustering in both real-space and in multipoles of redshift space. Estimates of the DESI statistical error are made using a jackknife method, allowing us to provide length-scales above which the halo clustering from our simulations is robust for DESI analysis.

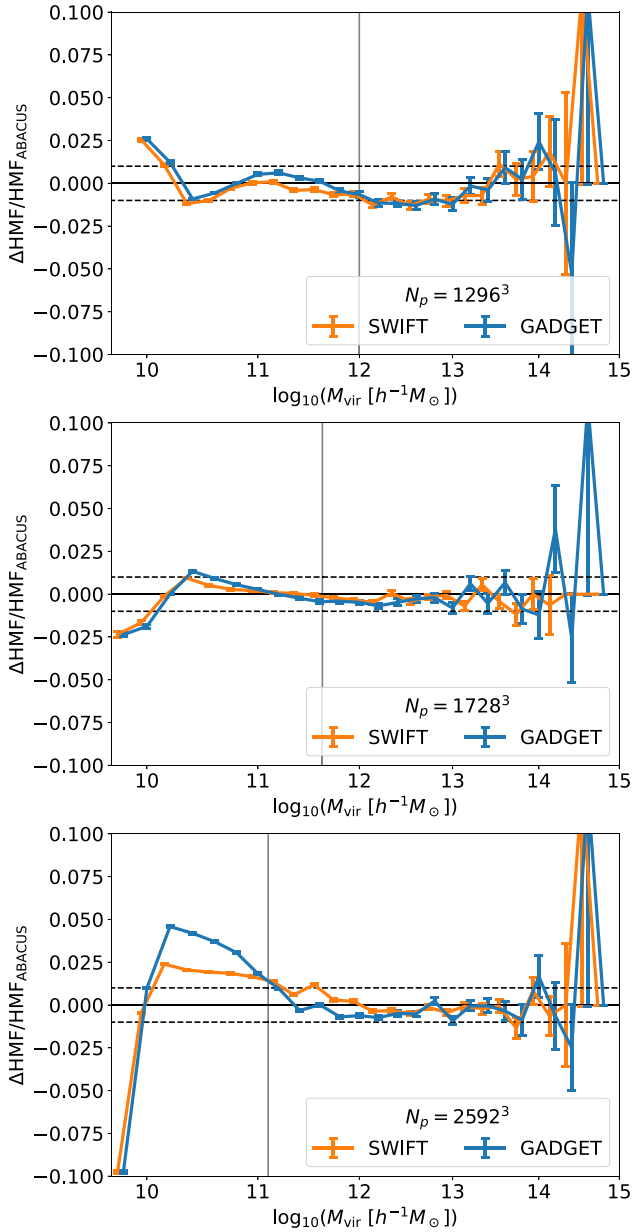

Figure 9. The ratio of the HMFs to that of the ABACUS simulation at fixed resolution at $z = 1$. The top, middle, and bottom panels show the low, medium, and high resolution simulations, respectively. The error bars show an estimate for the noise while comparing two simulations which share the same phases. Differences between the codes are below 1 per cent between 10^{12} and $10^{13.5}$ $h^{-1}\,\mathrm{M}_\odot$ at all resolutions. The differences in the HMF at large halo masses are consistent with noise. At low halo masses there is better agreement between the codes at low resolution than high resolution. Vertical lines indicate the mass at which haloes contain 200 particles at each resolution.

The two-point correlation function for haloes, $\xi(r)$ is defined as the excess probability of finding a halo at a distance r from another halo, averaged over all haloes.

$$\xi(|\mathbf{r}|) = \langle \delta(\mathbf{x})\delta(\mathbf{x}-\mathbf{r})\rangle. \quad (9)$$

The upper panel of Fig. 11 shows the real-space two-point correlation functions of the halo catalogues from the different codes relative to ABACUS. The high-resolution simulations were used with

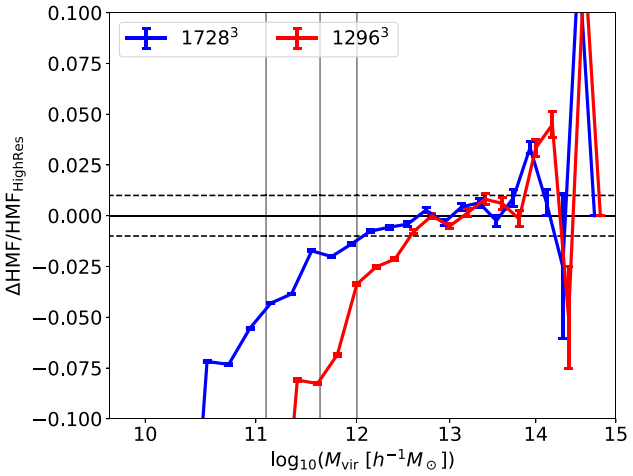

Figure 10. The ratio of the HMFs from the ABACUS simulations relative to the high-resolution run at $z = 1$. The error bars show an estimate for the noise while comparing two simulations which share the same phases. The 1296^3 simulation shows greater differences than 1728^3 to the 2592^3 simulation. There is a deficit in low-mass haloes in the lower resolution simulations relative to the high-resolution simulation. Vertical lines indicate the mass at which haloes contain 200 particles at each resolution.

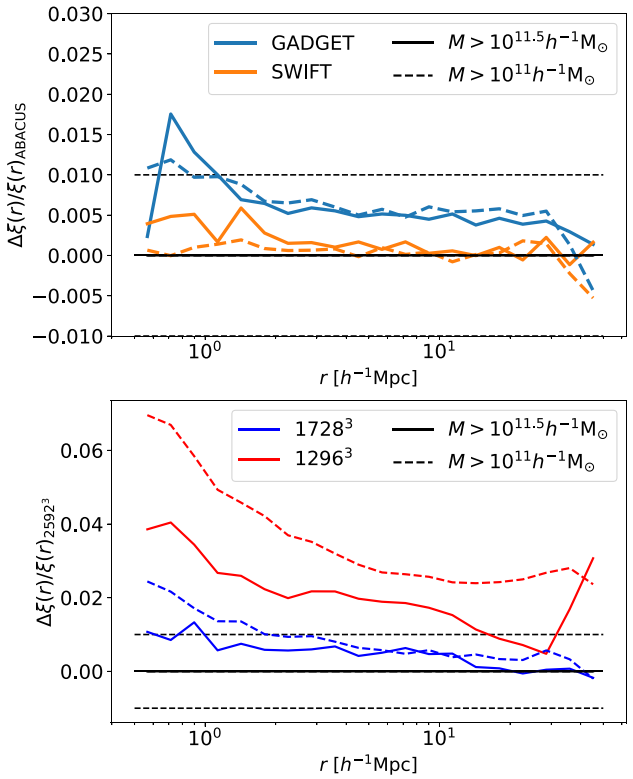

Figure 11. Upper panel: Real-space two-point correlation functions of the haloes from the 2592^3 simulations at $z = 1$. The results from using two halo mass limits are shown as solid and dashed lines. SWIFT and ABACUS agree more closely than GADGET-2 and ABACUS but for $r > 1$ Mpc all the codes agree to within 1 per cent. There are not significant differences which come from changing the halo mass limit. Lower panel: Real-space two-point correlation functions of the haloes from the ABACUS simulations relative to the high-resolution simulation at $z = 1$. Differences between the halo clustering measurements are larger at small scales. A higher halo mass limit and higher mass resolution both improve the agreement to the high-resolution two-point correlation function.

a mass cutoff $M > 10^{11.5}\ h^{-1}\,\mathrm{M}_\odot$ and $M > 10^{11}\ h^{-1}\,\mathrm{M}_\odot$ (solid and dashed lines, respectively). The correlation functions from the different simulation codes do not differ by greater than 1 per cent for length-scales above 1 h^{-1} Mpc.

In the lower panel of Fig. 11, we compare the real-space two-point correlation function of the haloes from ABACUS simulations at different resolutions. The high- and medium-resolution simulations agree to within 1 per cent for all length-scales above 1 h^{-1} Mpc. A larger difference is seen in the low-resolution halo clustering, the clustering increases at smaller scales to 4 per cent greater than the high resolution below 1 h^{-1} Mpc.

The halo catalogues from the low-resolution simulations have a higher clustering bias than the high-resolution case. This effect can be understood with reference to the HMF. (Fig. 10). There are fewer low-mass haloes in the low-resolution case, therefore the clustering measurement is made from a sample which has proportionally more high-mass haloes which have a higher bias (Mo & White 1996). Note that this clustering–resolution relation is in the opposite sense to that seen in DeRose et al. (2019) which saw lower bias at lower mass resolution. Differences in the treatment of force softening at different resolutions may contribute to this difference.

Decreasing the halo mass cut to $M > 10^{11}\ h^{-1}\,\mathrm{M}_\odot$ can be seen to reduce the agreement in the clustering relative to the higher mass cut. This is seen most strongly in the low-resolution simulations.

The multipoles of the redshift space correlation function, $\xi_l(s)$ are defined as (Peebles 1980; Kaiser 1987)

$$\xi_l(s) = \frac{2l+1}{2}\int_0^\pi \mathrm{d}\theta \sqrt{1-\mu^2}\xi(\sigma,\pi)\mathcal{L}_l(\mu). \tag{10}$$

Here, l is the multipole order, θ is the angle between the line of sight and the halo separation vector, $\mu = \cos(\theta)$, $\xi(\sigma, \pi)$ is the 2d correlation function, and $\mathcal{L}_l(\mu)$ is the l^{th} Legendre polynomial.

When making these comparisons of halo clustering, it is useful to understand the tolerance of observational measurements as this will provide a target accuracy. We estimate this observational tolerance by estimating the statistical error in the clustering using the jackknife method (Wu 1986; Norberg et al. 2009).

The ABACUS high-resolution box was used for jackknife estimates. This box was split into $N = 100$ distinct subvolumes. The redshift space correlation function multipoles were measured on N subsamples created by removing a single subvolume each time to provide N clustering estimates. The standard error of the clustering can then be found to be

$$\sigma(s) = \sqrt{\frac{N-1}{N}\sum_{i=0}^{N}\left(\xi^i(s) - \bar{\xi}(s)\right)^2}, \tag{11}$$

where $\xi^i(s)$ is the i^{th} clustering estimate, produced by removing the i^{th} subvolume and $\bar{\xi}(s)$ is the mean correlation function:

$$\bar{\xi}(s) = \sum_{i=0}^{N}\frac{\xi^i(s)}{N}. \tag{12}$$

The standard error estimate, $\sigma(s)$, corresponds to the statistical error for a cubic simulation box with side length 500 h^{-1} Mpc. For the full DESI survey volume we use an estimated volume of 20 h^{-3} Gpc^3 and for the year 1 DESI volume we use 4 h^{-3} Gpc^3 (Levi et al. 2013). Our estimate of the statistical error on the DESI survey is produced by multiplying by the square root of the volume ratio between the DESI survey and the simulation box:

$$\sigma_{\mathrm{DESI}}(s) = \sqrt{\frac{V_{\mathrm{sim}}}{V_{\mathrm{DESI}}}}\sigma(s). \tag{13}$$

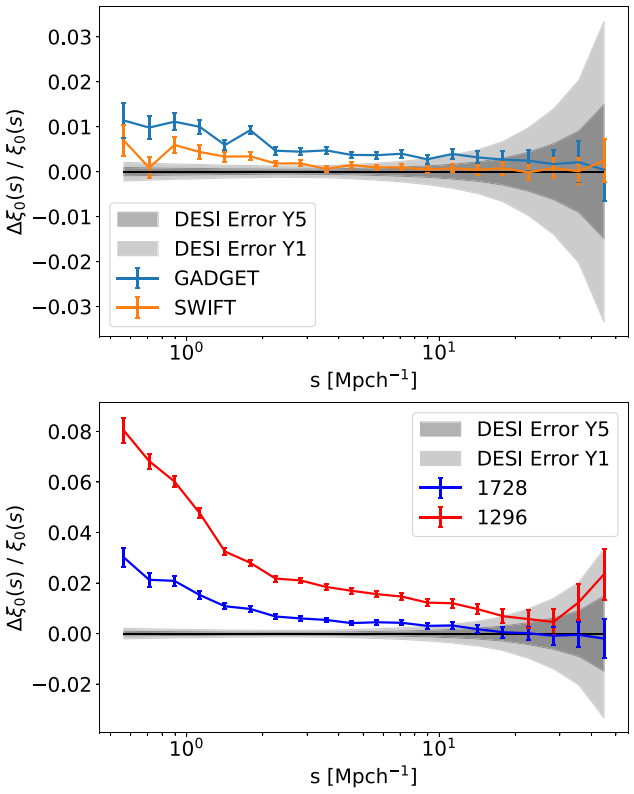

Figure 12. Upper panel: The redshift space clustering monopole comparison between the different codes at high resolution. ABACUS is used as the reference clustering. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The error bars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of $M > 10^{11.5}\ h^{-1}\,\mathrm{M}_\odot$ was used. The systematic differences between the codes are within the expected year 5 DESI volume statistical errors for $s > 20\ h^{-1}$ Mpc and within the year 1 volume statistical errors for $s > 10\ h^{-1}$ Mpc. Lower panel: As above but comparing the redshift space clustering between the different resolutions for the ABACUS code, with reference to the high-resolution simulation. The medium-resolution simulation is consistent with the high-resolution simulation to within the DESI statistical errors for $s > 20\ h^{-1}$ Mpc. The low-resolution clustering is not consistent with the high-resolution clustering to within the expected year 5 DESI statistical errors at any of the measured scales.

$\sigma_{\mathrm{DESI}}(s)$ is represented as a grey-shaded region on Figs 12–13.

This error estimate is likely to be tighter than the errors which are produced by DESI in practice. This is because over the full DESI volume the number density of observed galaxies will not be as high as the number density of haloes from the simulations used here. Modelling the galaxy–halo connection will be done in a future work (Hernández-Aguayo et al., in preparation).

A source of error in the clustering comparisons is that we are comparing the results from simulation boxes of finite size. The simulations were run from identical ICs, which removes sample variance. However our comparisons between different simulations with the same ICs still contain noise in the sense that the differences between these simulations contains a noise component that would change randomly if we changed the ICs and compared a second matched pair of simulations. This noise has the potential to obscure the measurement of systematic differences due to code and resolution. We use jackknife method to provide an estimate for this noise by using the variance in the clustering difference between matched

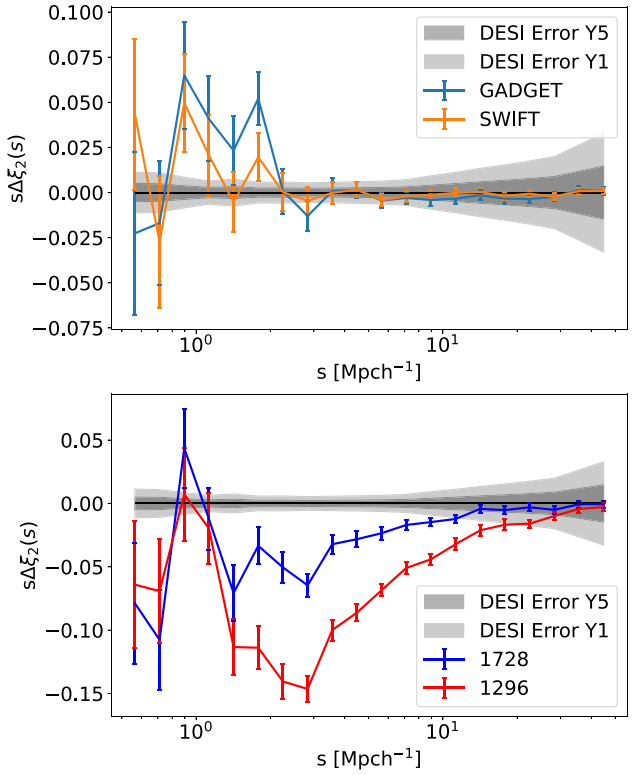

Figure 13. Upper panel: The redshift space clustering second-order multipole comparison between the different codes at high resolution. ABACUS is used as the reference clustering. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The error bars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of $M > 10^{11.5}\ h^{-1}\,\mathrm{M}_{\odot}$ was used. The codes agree to within the DESI statistical errors for $s > 10\,h^{-1}$ Mpc, below which noise begins to dominate the observed differences. Lower panel: As above but showing the redshift space clustering second-order multipole comparison between the different resolutions for the ABACUS code, with reference to the high-resolution simulation. There are significant differences compared to the level of the DESI statistical errors in both the low- and medium-resolution simulations for $s < 10\,h^{-1}$ Mpc.

subsamples:

$$\sigma_{\mathrm{NOISE}}(s) = \sqrt{\frac{N-1}{N}\sum_{i=0}^{N}\left(\Delta\xi^{i}(s) - \overline{\Delta\xi}(s)\right)^{2}}, \tag{14}$$

where $\Delta\xi^{i}(s)$ is the difference in the clustering measurements from two simulations with the i^{th} subvolume removed and $\overline{\Delta\xi}(s)$ is the mean clustering difference. A detailed investigation of the jackknife method estimating the noise between two simulations sharing the same phases will be presented in Zhang et al. (in preparation).

$\sigma_{\mathrm{NOISE}}(s)$ is represented as error bars on Figs 12–13. The clustering difference between ABACUS and each code is used for the code comparison noise estimate, while the clustering difference to the high-resolution ABACUS simulations is used for the resolution comparison noise estimate.

For the redshift space correlation function measurements, the results from three orthogonal lines of sight are averaged in order to reduce sample variance (Smith et al. 2020).

Figs 12 and 13 show the redshift space correlation function comparison between the high-resolution simulations for different codes in the monopole and quadrupole, respectively. Differences between the codes in the monopole are significant relative to the DESI statistical error below 10 Mpc h^{-1}, with sub 0.1 per cent agreement required for consistency within the DESI statistical errors at these length-scales. In the quadrupole the simulation noise dominates at small scales, making it difficult to establish systematic differences between the codes. The redshift space correlation function monopole and quadrupole difference between catalogues from different codes is smaller than the DESI statistical error above 20 h^{-1} Mpc. ABACUS and SWIFT have remarkable agreements down to 2 h^{-1} Mpc.

Figs 12 and 13 show the comparison between resolutions for the ABACUS simulations. The low-resolution correlation function monopole is not consistent with the high-resolution case to within the DESI statistical error at any length-scale. The medium resolution agrees with the high-resolution correlation function to within the DESI and simulation errors for $s > 20\,h^{-1}$ Mpc in the monopole and quadrupole.

One may naïvely expect that a threshold of N particles per halo is sufficient for a given clustering accuracy but that does not appear to be the case throughout these comparisons. In Fig 11, the medium-resolution simulation with a lower halo mass cut agrees with the high-resolution case more closely than the low-resolution simulation with a higher mass cut. This is likely to be because the force accuracy parameters of the simulations change with resolution and this could have effects at small length-scales.

Our results suggest that our medium- and high-resolution simulations have sufficient control over systematic errors in the redshift space correlation function for dark matter haloes in the regime where $s > 20\,h^{-1}$ Mpc for haloes with mass larger than $10^{11.5}\,h^{-1}\,\mathrm{M}_{\odot}$.

4.4 Halo clustering: power spectrum

The dark matter halo power spectrum is another useful metric to perform comparisons between N-body simulation codes. The definition of the real-space power spectrum can be found in Section 2.2.1. The multipole expansion of the redshift space power spectrum is used to understand the effects caused by peculiar velocities. The definition of the power spectrum multipole expansion is similar to the correlation function multipole expansion:

$$P_{l}(k) = (2l+1)\int_{0}^{1} \mathrm{d}\mu\, P(k,\mu)\mathcal{L}_{l}(\mu), \tag{15}$$

where l is the multipole order, μ is the cosine of the angle between the line of sight and the halo separation vector, $P(k, \mu)$ is the 2d, anisotropic power spectrum, and $\mathcal{L}_{l}$ is the l^{th} Legendre polynomial.

Figs 14–16 show the results from the halo power spectra comparisons between codes and resolutions, using two different halo mass cuts, in real-space and redshift space.

Jackknife errors are used to estimate the DESI survey statistical errors along with the noise due to finite simulation volume in a similar manner as in Section 4.3. A detailed investigation of the jackknife method estimating the uncertainties in the power spectrum measurements will be presented in Zhang et al. (in preparation).

In Fig. 14, the real-space power spectra are compared between codes. SWIFT, GADGET-2, and ABACUS agree at the 1 per cent level for $0.01 \leq k \leq 1\ h\,\mathrm{Mpc}^{-1}$. There is not a large difference between the power spectra for different halo mass cuts. The same effects are seen between the codes in the redshift space power spectrum monopole in Fig. 15, SWIFT, GADGET-2, and ABACUS are consistent within the DESI statistical errors up to $k = 0.2\ h\,\mathrm{Mpc}^{-1}$.

Fig. 14 shows that the medium- and low-resolution real-space halo power spectra are greater than the high-resolution case. The difference increases at smaller scales and is larger for the lower halo mass cut and for the lower resolution simulation. In Fig. 15, the same trends are seen in the redshift space monopole, with larger differences than

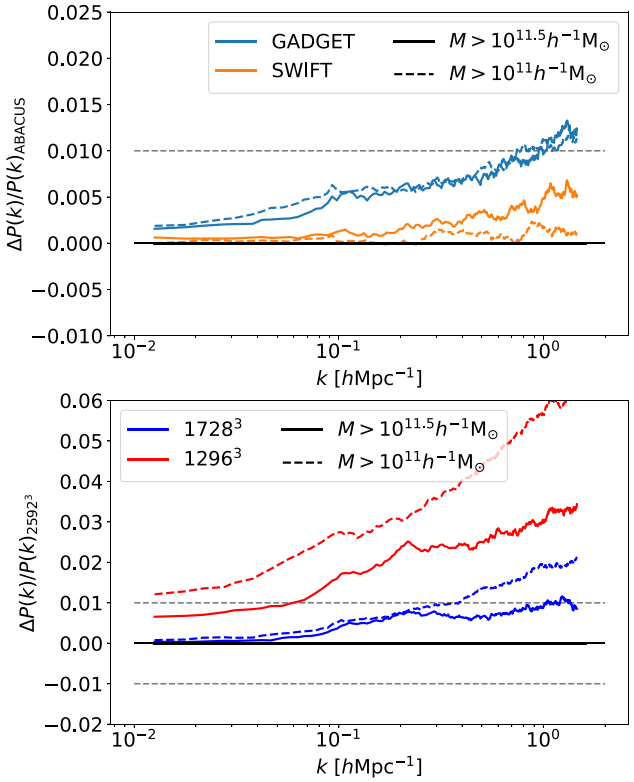

Figure 14. Upper panel: The ratio of the real-space power spectra of the haloes from the high-resolution simulations for different codes to ABACUS at $z = 1$ There is agreement between the codes to within the 1 per cent level at both mass cuts for $k < 1\ h\,\mathrm{Mpc}^{-1}$, with SWIFT and ABACUS showing agreement to within 0.1 per cent for $k < 0.1\ h\,\mathrm{Mpc}^{-1}$. Lower panel: The ratio of the real-space power spectra of the ABACUS haloes compared to the high-resolution halo catalogue. The medium-resolution simulation agrees with the high-resolution simulation more closely than the low resolution. The halo mass cut becomes more important for the resolution comparison than in the code comparison. The agreement becomes better with a higher halo mass cut in the resolution comparison.

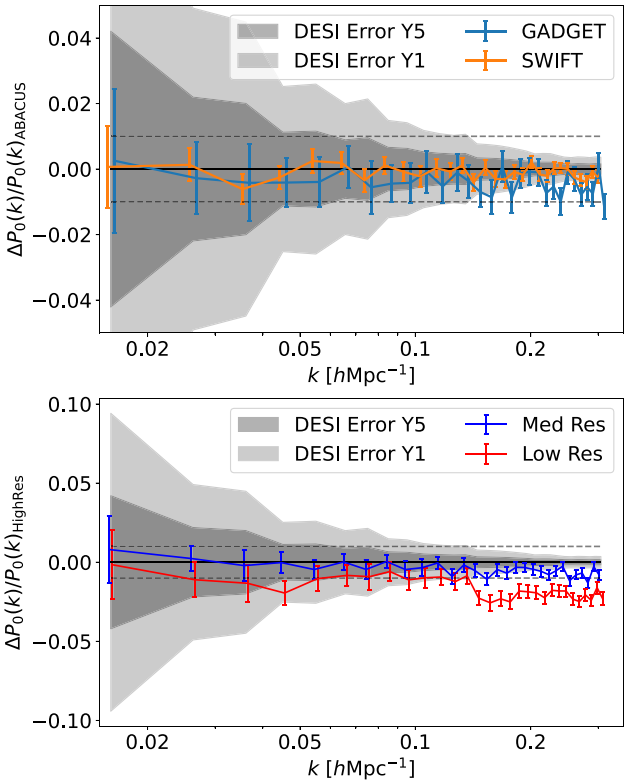

Figure 15. Upper panel: The ratio of the redshift space power spectrum monopoles of the haloes from the high-resolution simulations for different codes compared to ABACUS at $z = 1$. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The error bars show an estimate for the noise while comparing two simulations which share the same phases. The codes show agreement to within the expected DESI statistical errors for $k < 0.1\ h\,\mathrm{Mpc}^{-1}$. Lower panel: The ratio of the redshift space power spectrum monopoles from the high-resolution simulations for ABACUS at $z = 1$. The medium-resolution simulation is consistent with the high-resolution simulation for $k < 0.2\ h\,\mathrm{Mpc}^{-1}$. The low-resolution simulation shows significant differences for $k > 0.15\ h\,\mathrm{Mpc}^{-1}$.

the real-space power spectrum. The medium-resolution simulation is consistent with the high-resolution simulation for $k < 0.2\ h\,\mathrm{Mpc}^{-1}$. The low-resolution simulation shows significantly greater differences than the medium-resolution simulation for $k > 0.15\ h\,\mathrm{Mpc}^{-1}$.

The differences between the codes and resolutions in the redshift space quadrupole are shown in Fig. 16. Due to the level of noise it is difficult to draw conclusions about code and resolution agreement in the quadrupole. All the codes agree within the noise level, which is comparable to the DESI statistical error for $k < 0.05\ h\,\mathrm{Mpc}^{-1}$. At smaller scales than this the simulation noise becomes larger than the expected DESI statistical error.

5 CONCLUSIONS

We have performed two code comparison projects, comparing ICs generators and also *N*-body simulation codes.

5.1 IC comparison

The ICs code comparison included the IC generators associated with the ABACUS, and FASTPM *N*-body simulation codes, along with the IC generator PANPHASIA/IC_GEN. ICs were generated at three different resolutions for each code. Fig. 1 shows that the power spectra of the phase matched ICs from the same code agreed to within 1 per cent up to 20 per cent of the Nyquist frequency. At large scales the power spectra were consistent with the reference power spectrum, with any differences being within the expected variance due to finite volume.

At small scales, ABACUS, and PANPHASIA/IC_GEN agree with the input linear power spectrum within 1 per cent up to 50 per cent of the Nyquist frequency. Although we chose PANPHASIA/IC_GEN to set up the *N*-body code comparison project, the effect of the differences among the IC codes should be small.

The pairwise velocity anisotropy was used to check the consistency of the IC codes in the particle velocities as shown in Fig. 2. The codes were mostly consistent with one another to within the sample variance estimated by using 10 realizations of the PANPHASIA/IC_GEN ICs. FASTPM showed larger pairwise anisotropy than the other codes at large scales.

5.2 *N*-body simulation code comparison

The *N*-body simulation codes, ABACUS, SWIFT, GADGET-2, and FASTPM were run from identical PANPHASIA/IC_GEN ICs from $z = 199$ to $z = 0$. The matter power spectra were compared at $z = 2$ and

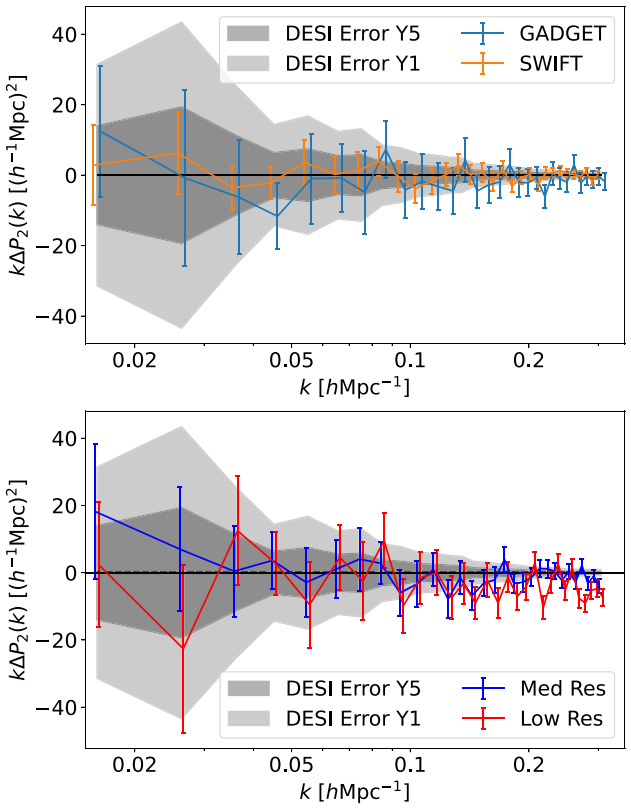

Figure 16. Upper panel: The difference of the redshift space power spectrum quadrupoles of the haloes from the high-resolution simulations for different codes compared to ABACUS at $z = 1$. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The error bars show an estimate for the noise while comparing two simulations which share the same phases. Lower panel: The difference of the redshift space power spectrum quadrupoles from the high-resolution simulations for ABACUS at $z = 1$. Due to the level of noise it is difficult to place bounds on the agreement of different codes and resolutions. For $k < 0.05\ h\,\mathrm{Mpc}^{-1}$, where the simulation noise is comparable in magnitude to the DESI errors, the code and resolution comparisons appear to be consistent with one another.

$z = 1$. Comparisons between codes at fixed resolution were made, along with comparisons between resolutions for the same code.

When comparing between resolutions for each code in Fig. 3, the power spectra agree to within 1 per cent at $k = 2\ h\ \mathrm{Mpc}^{-1}$ and to within 4 per cent at $k = 10\ h\ \mathrm{Mpc}^{-1}$ with the exception of the quasi-*N*-body code FASTPM which shows greater differences. Comparing between the codes at fixed resolution the power spectra agree to within 0.1 per cent at large scales as seen in Fig. 4. Fig. 5 shows that the large-scale power spectra are consistent with perturbation theory estimates made using 2LPT for all the codes. At small scales SWIFT, GADGET-2, and ABACUS show good agreement, within 1 per cent at $k = 10\ h\ \mathrm{Mpc}^{-1}$ for all resolutions at both redshifts. As expected, FASTPM shows the greatest difference in power spectrum relative to the other codes, disagreeing at the 1 per cent level for $k > 0.5\ h\ \mathrm{Mpc}^{-1}$. Similar conclusions are reached at the two explored redshifts.

The dark matter haloes were found using ROCKSTAR. Halo properties were compared between matched haloes for fixed code and variable resolution or fixed resolution and variable code and the results were shown in Figs 7 and 8. There was the greatest agreement between codes for high-mass haloes in most statistics. The difference between matched halo properties was within 1 per cent between all the codes for halo masses greater than $10^{13}\ h^{-1}\,\mathrm{M}_\odot$.

Matching halo properties between different resolutions presented several systematic effects. Low-mass haloes in low-resolution simulations had greater mass than the matched haloes from the high-resolution simulations, with the opposite effect observed at high masses. Matched halo properties had better agreement at high masses when comparing between simulations run with different resolutions.

The HMFs from SWIFT, GADGET-2, and ABACUS agreed to within 1 per cent between mass limits of $10^{11.5}$ and $10^{14}\ h^{-1}\,\mathrm{M}_\odot$ as seen in Fig. 9.

Halo clustering measurements were made by using both the correlation function and power spectrum. Comparisons were made between codes and resolutions for two different halo mass limits 10^{11} and $10^{11.5}\ h^{-1}\,\mathrm{M}_\odot$. The minimum halo mass that hosts the DESI emission line galaxy sample is expected to be between these two values.

Comparing different codes in the halo clustering, all codes agree with each other within the expected DESI year 1 uncertainty at scales larger than $10\ h^{-1}$ Mpc and within the DESI year 5 uncertainty at scales larger than $20\ h^{-1}$ Mpc (Figs 11–13). ABACUS and SWIFT have remarkable agreements down to $2\ h^{-1}$ Mpc. Comparing the different resolutions, the medium-resolution run agrees with the high-resolution one at scales larger than $20\ h^{-1}$ Mpc. This indicates that one should be more careful choosing simulation resolution than choosing the *N*-body code.

In the halo power spectra, similar results to the halo correlation function were found as seen in Figs 14–16. ABACUS, SWIFT, and GADGET-2 have good agreement at all the scales which we are most interested in for DESI, i.e. $k < 0.3\ h\ \mathrm{Mpc}^{-1}$. Medium-resolution simulations agree with the high-resolution simulations at a similar level to the code comparisons. The low-resolution simulation shows differences at least twice as large as the medium-resolution simulation at all length-scales. The agreement becomes closer between different resolution simulations with a higher halo mass cut.

These results indicate the expected level of systematic errors in a variety of statistics associated with *N*-body simulations.

We do not draw conclusions about whether certain simulations are appropriate for use in any specific DESI analyses, this requires further propagation of the simulations through the cosmology analysis pipeline. These simulations and the results provided should act as a baseline for expected systematic errors, regardless of the application of the simulations. Future work using these simulations will explore how these errors are propagated through halo occupation distribution and cosmological analysis and therefore will assess the suitability of simulations run with different codes and resolutions for use in modern galaxy surveys (Hernández-Aguayo et al., in preparation).

ACKNOWLEDGEMENTS

We thank Violeta Gonzalez-Perez, Graziano Rossi, Peter Behroozi, Joe DeRose (internal reviewer), and Sownak Bose (internal reviewer) for their constructive comments. This research is supported by the Director, Office of Science, Office of High Energy Physics of the US Department of Energy under Contract No. DE-AC02-05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technology Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation;

the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology, Mexico (CONACYT); the Ministry of Science and Innovation of Spain (MICINN), and by the DESI Member Institutions: https://www.desi.lbl.gov/collaborating-institutions.

The authors are honoured to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation.

CG is supported by a PhD Studentship from the Durham Centre for Doctoral Training in Data Intensive Science, funded by the UK Science and Technology Facilities Council (STFC, ST/P006744/1) and Durham University. LHG is supported by the Center for Computational Astrophysics at the Flatiron Institute, which is supported by the Simons Foundation. ABACUS development has been supported by NSF AST-1313285 and DOE-SC0013718, as well as by Harvard University startup funds. BL would like to acknowledge the support of the National Research Foundation of Korea (NRF-2019R1I1A1A01063740) and the support of the Korea Institute for Advanced Study (KIAS) grant funded by the government of Korea. CH-A acknowledges support from the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC-2094-390783311. This work used the DiRAC@Durham facility managed by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). The equipment was funded by BEIS capital funding via STFC capital grants ST/K00042X/1, ST/P002293/1, ST/R002371/1, and ST/S002502/1, Durham University and STFC operations grant ST/R000832/1. DiRAC is part of the National e-Infrastructure. The research in this paper made use of the SWIFT open-source simulation code (http://www.swiftsim.com, Schaller et al. 2018) version 0.9.0.

DATA AVAILABILITY

The authors make the ICs files, simulation snapshots, and halo catalogues available at a Globus link.[11] We encourage their use as a valuable resource for comparing the impact of *N*-body simulation code and resolution in other projects.

REFERENCES

Abrahamsen P., 1997, A Review of Gaussian Random Fields and Correlation Functions. Norsk Regnesentral/Norwegian Computing Center, Oslo
Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2013, ApJ, 762, 109
Benson A. J., Cole S., Frenk C. S., Baugh C. M., Lacey C. G., 2000, MNRAS, 311, 793
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, MNRAS, 300, 945
Conroy C., Wechsler R. H., 2009, ApJ, 696, 620
Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dehnen W., 2014, Comput. Astrophys. Cosmol., 1, 1
DeRose J. et al., 2019, ApJ, 875, 69
DeRose J. et al., 2022, Phys. Rev. D, 105, 123520
DESI, 2016, preprint (arXiv:1611.00036)
Eke V. R., Cole S., Frenk C. S., 1996, MNRAS, 282, 263
Feng Y., Chu M.-Y., Seljak U., McDonald P., 2016, MNRAS, 463, 2273
Garrison L. H., Eisenstein D. J., Ferrer D., Metchnik M. V., Pinto P. A., 2016, MNRAS, 461, 4125
Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, MNRAS, 485, 3370
Garrison L. H., Joyce M., Eisenstein D. J., 2021a, MNRAS, 504, 3550
Garrison L. H., Eisenstein D. J., Ferrer D., Maksimova N. A., Pinto P. A., 2021b, MNRAS, 508, 575
Hand N., Feng Y., Beutler F., Li Y., Modi C., Seljak U., Slepian Z., 2018, AJ, 156, 160
Jenkins A., 2010, MNRAS, 403, 1859
Jenkins A., 2013, MNRAS, 434, 2094
Jong J., Kleijn G. A. V., Kuijken K. H., Valentijn E. A., 2012, Exp. Astron., 35, 25
Kaiser N., 1987, MNRAS, 227, 1
Levi M. et al., 2013, preprint (arXiv:1308.0847)
Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, MNRAS, 508, 4017
Marcos B., Baertschiger T., Joyce M., Gabrielli A., Labini F. S., 2006, Phys. Rev. D, 73, 103507
Mo H. J., White S. D. M., 1996, MNRAS, 282, 347
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neyrinck M. C., Yang L. F., 2013, MNRAS, 433, 1628
Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, MNRAS, 396, 19
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Peebles P. J. E., 1980, The Large-Scale Structure of the Universe. Princeton Univ. Press, Princeton, NJ
Planck Collaboration VI, 2020, A&A, 641, A6
Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14
Rodríguez-Torres S. A. et al., 2016, MNRAS, 460, 1173
Schaller M., Gonnet P., Chalk A. B. G., Draper P. W., 2016, Proc. Platform Adv. Sci. Comput. Conf. PASC'16. Association for Computing Machinery, NY, USA, p. 1 https://dl.acm.org/doi/proceedings/10.1145/2929908
Schaller M., Gonnet P., Draper P. W., Chalk A. B. G., Bower R. G., Willis J., Hausammann L., 2018, Astrophysics Source Code Library, record ascl:1805.020
Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 2016, 047
Smith A., de Mattia A., Burtin E., Chuang C.-H., Zhao C., 2020, MNRAS, 500, 259
Springel V., 2005, MNRAS, 364, 1105
Springel V. et al., 2005, Nature, 435, 629
Theuns T., Chalk A., Schaller M., Gonnet P., 2015, preprint (arXiv:1508.00115)
Wechsler R. H., Tinker J. L., 2018, ARA&A, 56, 435
Winther H. A. et al., 2015, MNRAS, 454, 4208
Wu C. F. J., 1986, Ann. Stat., 14, 1261
Zel'Dovich Y. B., 1970, A&A, 500, 13

APPENDIX A: COMPARISON OF SUB-HMFS

The comparison of sub-HMFs between the simulations at fixed resolution at $z = 1$ is presented in Fig. A1. The differences between the sub-HMFs are smaller at lower resolution, lying within 1, 2, and 5 per cent in the low, medium, and high resolutions, respectively between 10^{11} and10^{12} $h^{-1}\,\mathrm{M}_{\odot}$. SWIFT and GADGET-2 show good agreement within 1 per cent for all resolutions between the mass ranges of 10^{11} and 10^{12} $h^{-1}\,\mathrm{M}_{\odot}$. Above 10^{13} $h^{-1}\,\mathrm{M}_{\odot}$, the sub-HMFs become noisy. The sub-HMF is likely to be more sensitive to differences in the softening schemes of the different simulations than the parent HMF.

[11] https://app.globus.org/file-manager?origin_id=53feea72-5215-11ec-a9cc-91e0e7641750&origin_path=~per~cent2F

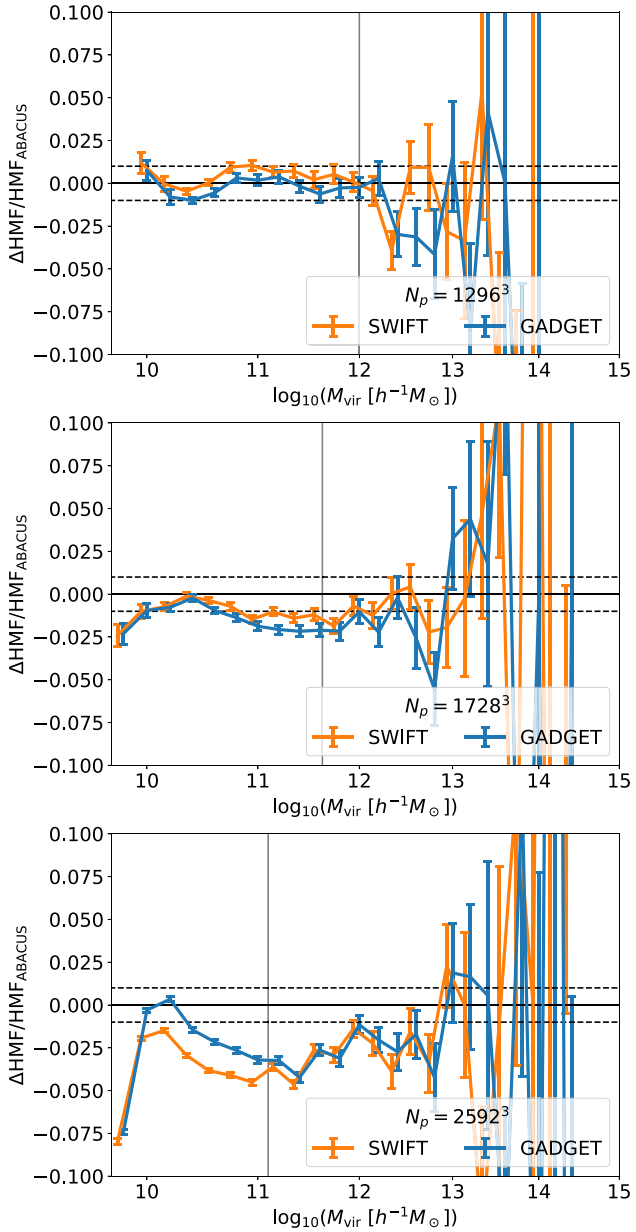

Figure A1. The ratio of the sub-HMFs to that of the ABACUS simulation at fixed resolution at $z = 1$. The top, middle, and bottom panels show the low-, medium-, and high-resolution simulations, respectively. The error bars show an estimate for the noise while comparing two simulations which share the same phases. Vertical lines indicate the mass at which haloes contain 200 particles at each resolution.

[1]*Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK*
[2]*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*
[3]*Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA*
[4]*Instituto de Astronomia, Universidad Nacional Autónoma de México, Apartado Postal 20-364, México*
[5]*Department of Physics, Manipur University, Canchipur, Manipur 795003, India*
[6]*Center for Computational Astrophysics, Flatiron Institute Simons Foundation, 162 Fifth Avenue, New York, NY 10010, USA*
[7]*Department of Physics and Astronomy, Sejong University, Seoul 05006, Korea*
[8]*Berkeley Center for Cosmological Physics, Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA*
[9]*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany*
[10]*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*
[11]*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[12]*Department of Physics, Kansas State University, Manhattan, KS 66506, USA*
[13]*Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China*
[14]*Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA*
[15]*Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham DH1 3LE, UK*
[16]*University College London, Department of Physics & Astronomy, Gower Street, London WC1E 6BT, UK*
[17]*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*
[18]*NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N Cherry Ave, Tucson, AZ 85719, USA*
[19]*Space Sciences Laboratory (SSL), University of California Berkeley, CA 94720, USA*
[20]*Physics Department, University of Michigan Ann Arbor, MI 48109, USA*

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