

# Lawrence Berkeley National Laboratory

LBL Publications

## Title

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

## Permalink

<https://escholarship.org/uc/item/9vp5x16c>

## Journal

Monthly Notices of the Royal Astronomical Society, 515(2)

## ISSN

0035-8711

## Authors

Grove, Cameron

Chuang, Chia-Hsun

Devi, Ningombam Chandrachani

et al.

## Publication Date

2022-07-26

## DOI

10.1093/mnras/stac1947

## Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at <https://creativecommons.org/licenses/by-nc/4.0/>

Peer reviewed

# The DESI $N$ -body Simulation Project I: Testing the Robustness of Simulations for the DESI Dark Time Survey

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

Affiliations are listed at the end of the paper

Accepted XXX. Received YYY; in original form ZZZ

## ABSTRACT

Analysis of large galaxy surveys requires confidence in the robustness of numerical simulation methods. The simulations are used to construct mock galaxy catalogs 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,  $6.25 \times 10^8$ ,  $2.11 \times 10^9$ , and  $5.00 \times 10^9 h^{-1} M_{\odot}$ , matching phases to reduce the noise within the comparisons. We find systematic errors in the halo clustering between different codes are smaller than the DESI statistical error for  $s > 20 h^{-1} \text{Mpc}$  in the correlation function in redshift space. Through the resolution comparison we find that simulations run with a mass resolution of  $2.1 \times 10^9 h^{-1} M_{\odot}$  are sufficiently converged for systematic effects in the halo clustering to be smaller than the DESI statistical error at scales larger than  $20 h^{-1} \text{Mpc}$ . 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% for  $k \leq 10 h \text{Mpc}^{-1}$ . 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:** large-scale structure of Universe – software: simulations – galaxies: haloes – cosmology: theory – methods: numerical

## 1 INTRODUCTION

For many years  $N$ -body simulations have been used as a tool to explore the nonlinear evolution of the distribution of matter in the Universe (Davis et al. 1985). Their use has been invaluable in creating mock galaxy catalogs to validate the results of surveys such as 2dFGRS, SDSS, DES, KiDS, and more (Cole et al. 1998; Rodríguez-Torres et al. 2016; DeRose et al. 2021; de Jong et al. 2012). The Dark Energy Spectroscopic Instrument (DESI) will be performing surveys of unprecedented size over the next five years, 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 catalogs 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

\* E-mail: cameron.grove@durham.ac.uk

† E-mail: albert.chuang@utah.edu

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. 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 et al. 2019).

This paper contains the results from multiple code comparison projects using several modern  $N$ -body simulation methods. Firstly we compare the initial conditions 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 ABACUS, FASTPM, GADGET-2, and SWIFT from identical initial conditions. 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 initial conditions 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 catalogs produced from the different simulation codes, including measurements of the clustering, halo mass functions, 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 initial conditions (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 et al. 2020). The properties of this field are determined by its power spectrum, which in turn is determined by the cosmology of the universe. Different realisations 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<sup>1</sup> (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  $\Lambda$ CDM Planck 2018 results (Planck Collaboration et al. 2020) with the following parameters: the Hubble expansion rate at present time,  $H_0 = 67.36 \text{ km s}^{-1}$ ,  $\Omega_c h^2 = 0.1200$ ,  $\Omega_b h^2 = 0.0223$ ,  $\Omega_\Lambda = 0.6834$ , r.m.s linear density fluctuation,  $\sigma_8 = 0.8$ , scalar spectral index  $n_s = 0.9649$ .

Considering the requirements of the DESI project, we run the simulations in boxes of side length,  $L_{\text{box}} = 500h^{-1}\text{Mpc}$  with three different dark matter particle numbers:  $N_p = 1296^3$ ,  $1728^3$  and  $2592^3$ . The initial conditions are generated at redshift,  $z_{\text{ini}} = 199$  with the input theoretical matter power spectrum generated using CAMB<sup>2</sup>. The details of the methods employed by each code

**Table 1.** The number of particles and particle masses of the simulation boxes. All are cubic boxes with a side length of 500 Mpc/h.

Simulation	Particle Number	Particle Mass $/h^{-1}M_\odot$
Low Resolution	$1296^3$	$5.00 \times 10^9$
Medium Resolution	$1728^3$	$2.11 \times 10^9$
High Resolution	$2592^3$	$6.25 \times 10^8$

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 initial conditions 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<sup>3</sup>, generates Zel'dovich Approximation initial conditions, 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 initial conditions in memory-limited environments by buffering the state to disk. 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 initial conditions generator used by the ABACUS code, including the AbacusSummit project<sup>4</sup> (Maksimova et al. 2021).

#### 2.1.2 IC\_GEN / PANPHASIA

There are two components to this method of creating the initial conditions which has been labelled PANPHASIA. Firstly PANPHASIA describes a method of setting the phases of cosmological initial conditions. PANPHASIA describes a pseudo random number sequence mapped to a cosmological volume from which subsections can be taken to provide the phases for initial conditions (Jenkins 2013). This allows simple creation of resimulation initial conditions 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 initial conditions are generated using a proprietary code called IC\_GEN. This uses second order Lagrangian perturbation theory

<sup>1</sup> <http://icc.dur.ac.uk/Panphasia.php>

<sup>2</sup> <https://camb.info/>

<sup>3</sup> <https://github.com/abacusorg/zeldovich-PLT>

<sup>4</sup> <https://abacussummit.readthedocs.io>

(2LPT) (Jenkins 2010) to create the initial conditions particle distribution according to the phases specified by the location within PANPHASIA.

### 2.1.3 FastPM IC

While not originally designed as an initial condition code, FastPM does compute an initial particle state with the traditional 2LPT-on-a-mesh approach. In addition, the initial condition 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 and second order LPT 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  $P_{\text{lin}}(k)$ . The white noise field is sampled on a finite mesh of  $B_{\text{LPT}}N_c$  samples per side from a pseudo-random generator compatible to the one used by Gadget's N-GenIC (Angulo et al. 2012; Springel et al. 2005).

The choice of  $B_{\text{LPT}}$ , 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 Initial Conditions

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

### 2.2.1 Comparison of Particle Distribution

The power spectrum is a common measure for analysing both initial conditions 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,  $\delta(\mathbf{x}) = \frac{\rho(\mathbf{x})}{\bar{\rho}} - 1$ ,

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

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

Initial conditions 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 initial conditions. This is the Nyquist frequency. For a box with sidelength  $L$  and  $N^3$  particles the wavenumber associated with the Nyquist frequency is

$$k_{\text{Nyquist}} = \frac{\pi N}{L}. \quad (3)$$

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`<sup>5</sup> (Hand et al. 2018) using a triangular shaped cloud mass assignment scheme and grid sizes of  $2592^3$ ,  $3456^3$   $5184^3$  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 modes in a particular realisation 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% in all resolutions up to the Nyquist frequency. The FASTPM ICs have much lower power than the reference power spectrum for  $k > k_{\text{Nyquist}}/2$ . The ABACUS ICs have the best agreement with the reference power spectrum.

It should be noted that it is not imperative for initial conditions 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  $k_{\text{Nyquist}}$  are mostly unimportant to the growth of structure at these scales.

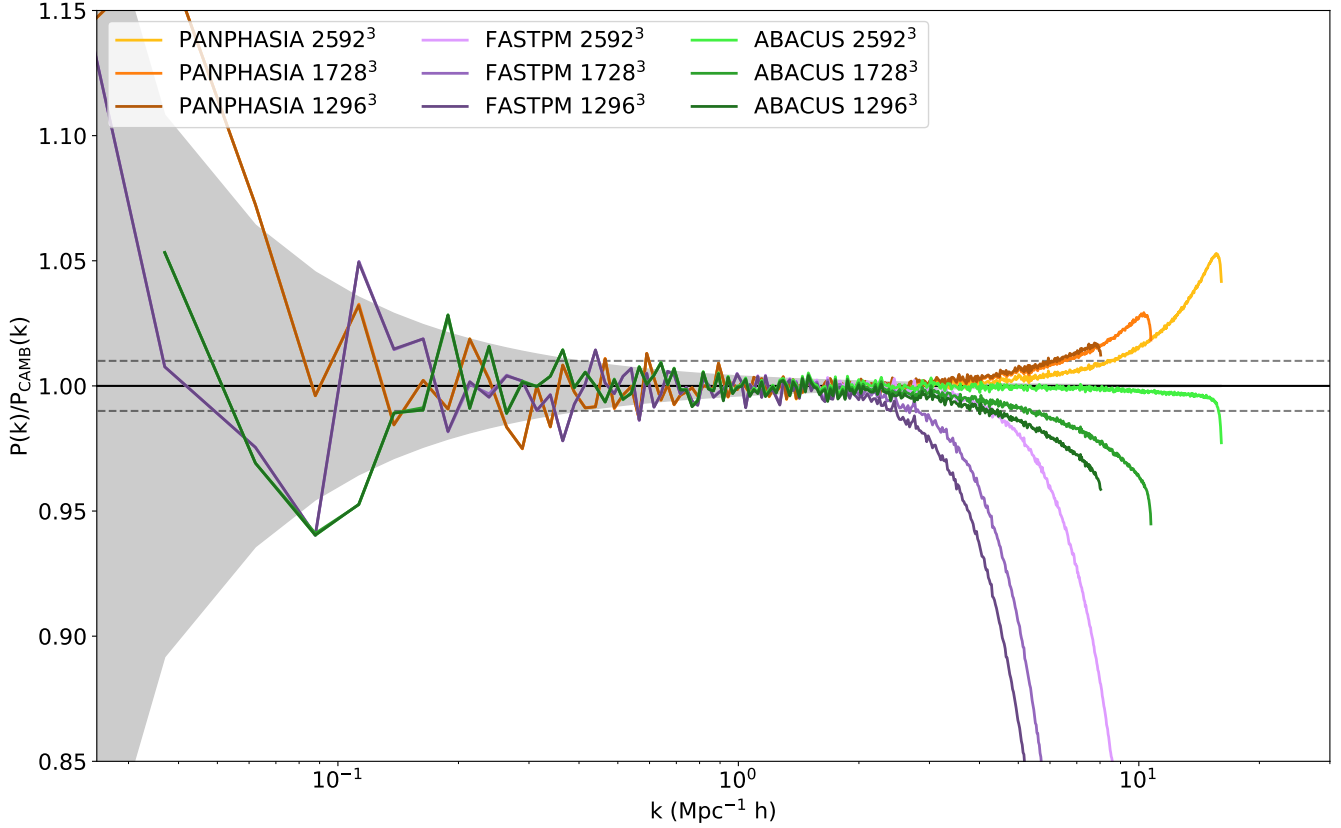
### 2.2.2 Comparison of Velocity statistics

The power spectrum is only a measure of the displacements of the particles produced by an initial conditions 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 initial conditions, such as the Zel'dovich approximation (Zel'dovich 1970) and 2LPT (Crocce et al. 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 realisations 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 realisations.

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  $(\mathbf{x}_2, \mathbf{v}_2)$

<sup>5</sup> <https://nbodykit.readthedocs.io/en/latest/>



**Figure 1.** A comparison of the power spectrum of initial conditions 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 1-sigma sample variance. The dashed lines indicate  $\pm 1\%$ .

the parallel and perpendicular components are

$$v_{\parallel} = \frac{(\mathbf{x}_1 - \mathbf{x}_2) \cdot (\mathbf{v}_1 - \mathbf{v}_2)}{|\mathbf{x}_1 - \mathbf{x}_2|} \quad (4)$$

and

$$v_{\perp} = \sqrt{|\mathbf{v}_1 - \mathbf{v}_2|^2 - v_{\parallel}^2} \quad (5)$$

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

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

and

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

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

The velocity anisotropy parameter

$$\beta = 1 - \frac{\sigma_{\perp}^2}{2\sigma_{\parallel}^2} \quad (8)$$

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

Fig. 2 shows that the velocity anisotropy, as a function of pairwise inter-particle distance, is mostly consistent between the codes and at large scales the differences are mostly explained by

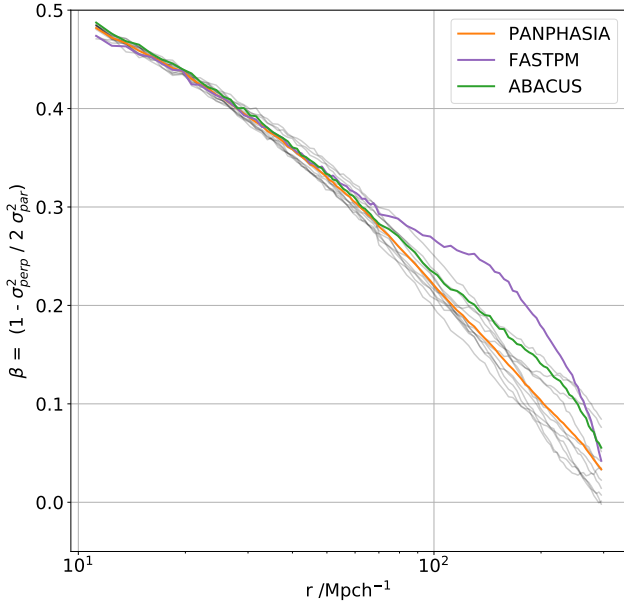
sample variance. There is excellent agreement between the codes at small inter-particle distance. FastPM is the only outlier, having a slightly higher anisotropy than the other realisations at around  $r = 150 h^{-1} \text{Mpc}$ .

It was decided to use PANPHASIA initial conditions 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 initial conditions codes in the future.

### 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 initial conditions 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 de-



**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 realisations 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 realisations around  $150 h^{-1}$  Mpc.

scribed 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  $10^{-5}$  to  $10^{-6}$ . 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 et al. (2021a). The multipole order was  $p = 8$ , with  $405^3$  cells, and the time step parameter was set to  $\eta_{\text{Accel}} = 0.25$ . This is mildly more conservative than ABACUS's run of the Euclid code comparison simulation, which used  $\eta_{\text{Accel}} = 0.3$ . Garrison et al. (2021b) show that a time step value of

$\eta = 0.25$  produces sub-percent 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 timesteps are evenly spaced in  $\log(a)$  with dynamic shorter timesteps 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 timestepping and force accuracy were  $\text{MaxRMSDisplacementFac} = 0.2$ ,  $\text{MaxSizeTimestep} = 0.01$ ,  $\text{ErrTolTheta} = 0.5$ , and  $\text{ErrTolForceAcc} = 0.0025$ .

#### 3.1.3 SWIFT

SWIFT<sup>7</sup> (Schaller et al. 2016) is a hydrodynamics and gravity code which uses a task based parallelisation 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 timesteps are evenly spaced in  $\log(a)$  and particles can have smaller timesteps if their acceleration exceeds a threshold (Theuns et al. 2015)<sup>8</sup>. 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<sup>9</sup> (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

<sup>6</sup> The Panphasia descriptor describing the simulation volume is [Panph1,L21,(1133107,236124,673886),S81,CH1311607586,DESI\_IC\_v1]

<sup>7</sup> <http://swift.dur.ac.uk/>

<sup>8</sup> <https://gitlab.cosma.dur.ac.uk/swift/swiftsim>

<sup>9</sup> <https://github.com/rainwoodman/fastpm>



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. \(2021\)](#).

### 3.2 Comparison of particle distribution

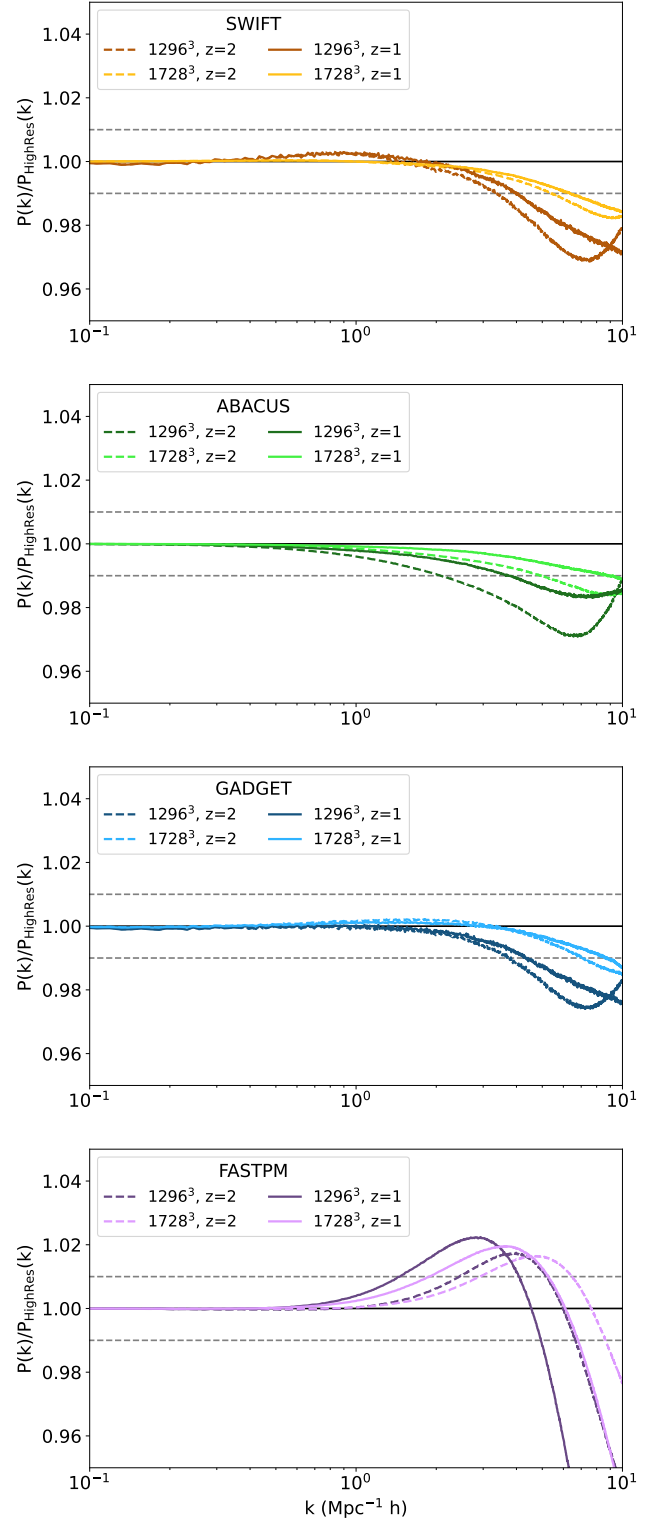
We first begin by examining the power spectrum dependence on the mass resolution for each code separately. As seen in Fig 3, this dependence is similar for all the codes tested. There is very little change in the power spectra at large scales and at small scales there are differences of 1-2%. The higher 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% at  $k = 2 h \text{ Mpc}^{-1}$ . 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  $k = 10 h \text{ Mpc}^{-1}$  the fractional differences between the power spectra are below 1%.

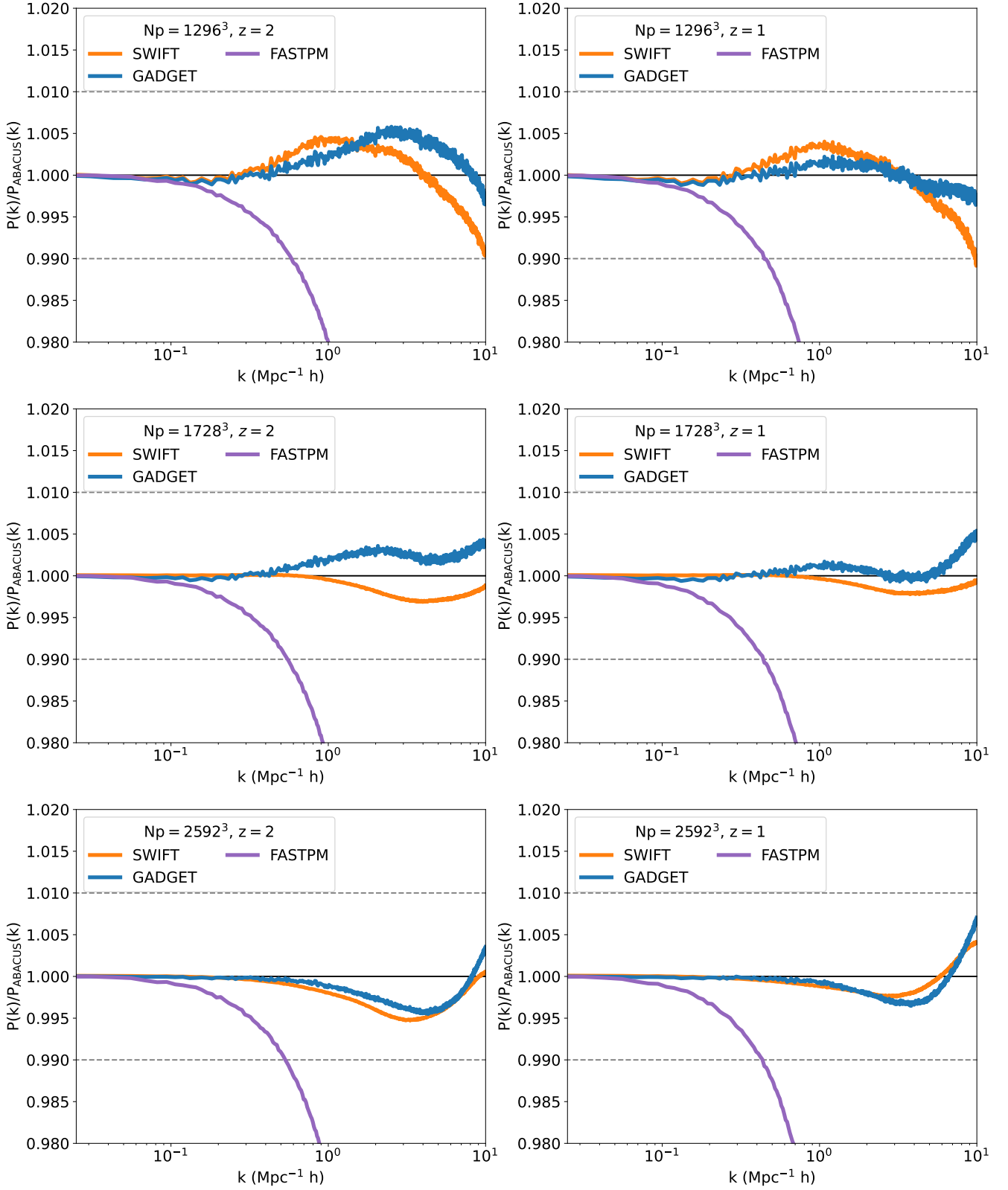
All codes agree on the low- $k$  power spectrum amplitude to within 0.1%, but to ascertain that this value is that predicted by perturbation theory, the 2LPT estimates of the  $z = 1$  and  $z = 2$  power spectra were measured by using IC\_GEN, the code used to create the initial conditions for the simulations. 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 Figure 5. We also attempted to compare to linear theory by rescaling the power spectrum of the initial conditions by the linear growth factor, however this did not reproduce the large scale power spectrum to sub-percent accuracy, as shown in the dashed lines in Figure 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% 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 over-dense regions called dark matter halos because these are the sites that host galaxies ([Peacock & Smith 2000](#); [Benson et al. 2000](#); [Conroy & Wechsler 2009](#)). Using and storing full particle data can become prohibitively expensive for large  $N$ -body simulations therefore halo catalogs are often the most useful data product to be produced from an  $N$ -body simulation. Observations and simulations suggest galaxies reside within dark matter halos and therefore they are essential

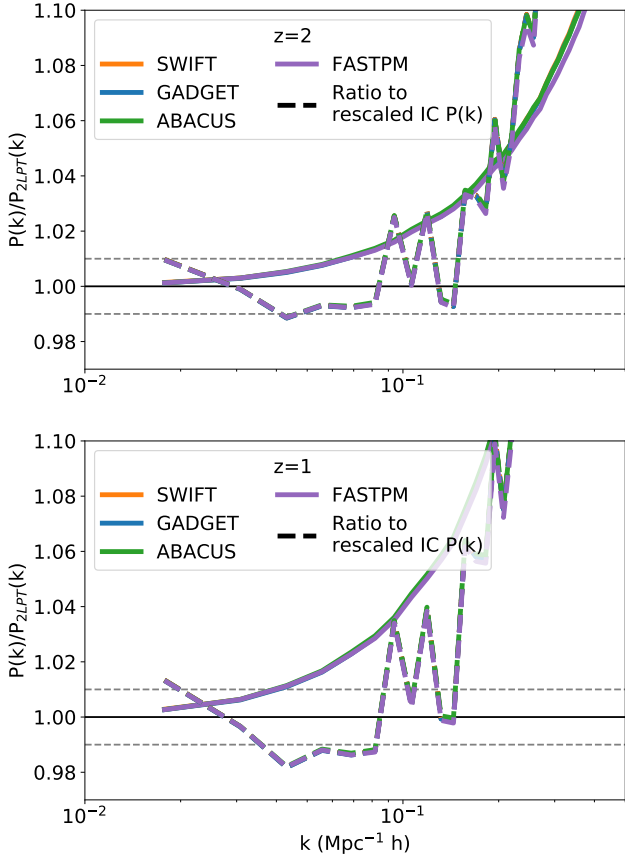


**Figure 3.** The matter power spectrum of the low and medium resolution runs relative to that of the high resolution run for each code at  $z = 1$  and 2. SWIFT, ABACUS, & GADGET-2 show similar behaviour where the power spectrum decreases by up to 2% at  $k = 10 h \text{ Mpc}^{-1}$  when decreasing 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  $k = 3 h \text{ Mpc}^{-1}$  before decreasing sharply above  $k = 5 h \text{ Mpc}^{-1}$ .



**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% at  $k < 0.1 h \text{ Mpc}^{-1}$ . Differences emerge at high  $k$  but, excepting FASTPM, these are within 1% at  $k = 10 h \text{ Mpc}^{-1}$ .





**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 initial conditions rescaled by the growth factor which do not agree as closely as 2LPT.

to connect  $N$ -body simulations to observable results (Wechsler & Tinker 2018).

Halo catalogs can be used to create mock galaxy catalogs 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); DeRose et al. (2021).

In this section we perform the comparison of halo properties, mass functions and clustering. Halos are identified with the help of the phase-space temporal halo finder ROCKSTAR with force resolution parameters shown in Table 2 (Behroozi et al. 2013)<sup>10</sup>, and default parameters otherwise, including STRICT\_SO\_MASSES=0.

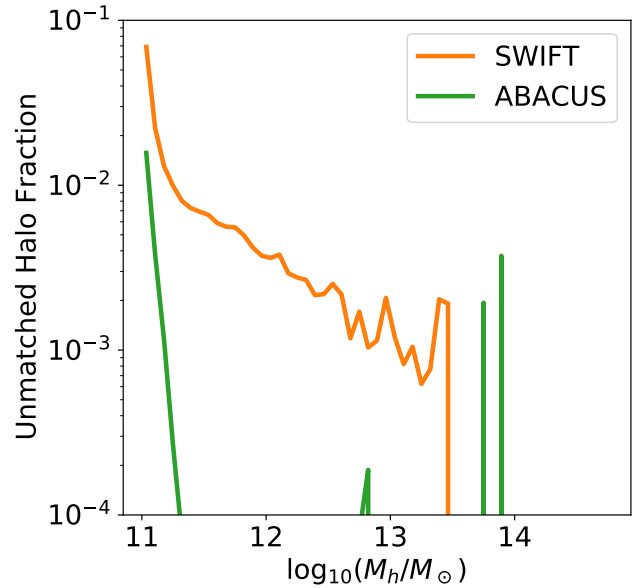
We do not include FASTPM in this comparison of halo properties and clustering. This is because performing abundance matching is outside the scope of this paper, and the FASTPM results without abundance matching would be unrealistically discrepant compared with how the code is used in practice.

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

<sup>10</sup> <https://bitbucket.org/gfcstanford/rockstar>

**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. Halos 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



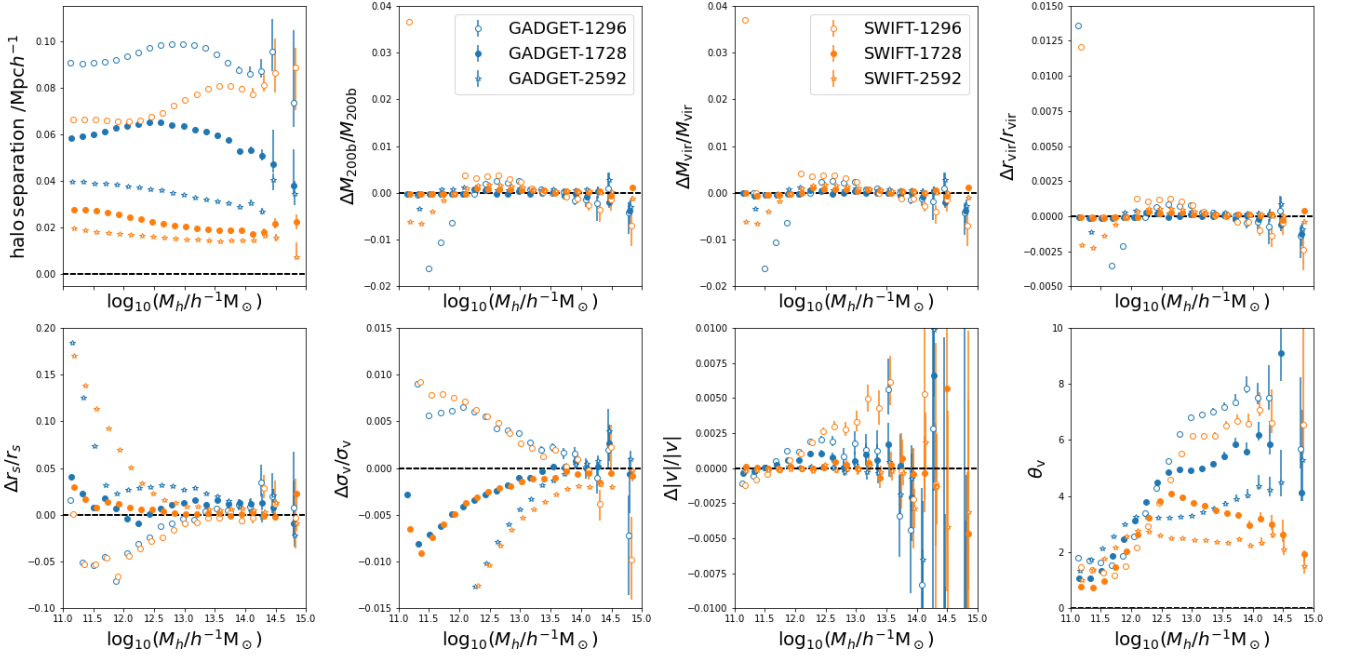
**Figure 6.** The fraction of halos which remain unmatched as a function of halo mass in the ABACUS and SWIFT catalogs when matching halos between the high resolution catalogs. The majority of halos are matched across simulations.

#### 4.1 Halo properties

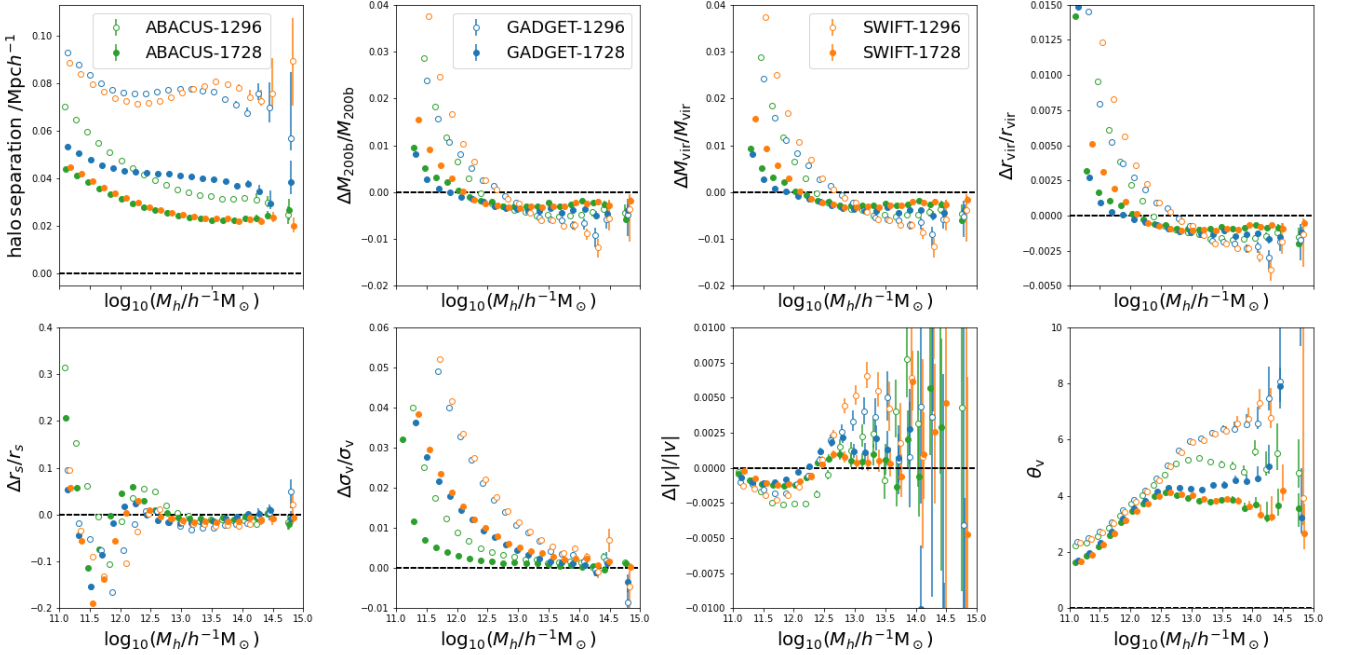
We execute a detailed halo-to-halo comparison by matching halos between  $N$ -body simulations and comparing their properties. There is no guarantee that halos will form in identical places, but since all simulations share the same initial particles, in many cases a robust match can be identified. Still,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

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

Figure 6 shows the fraction of unmatched halos in each catalog in a comparison between the high resolution SWIFT and ABACUS simulations. This fraction increases as halo mass decreases but is only greater than 5% of halos for the lowest mass SWIFT halos at  $M_h = 10^{11} h^{-1} M_{\odot}$ . This gives us confidence that there are



**Figure 7.** A comparison of the halo properties for halos matched between ABACUS and the other codes at fixed resolution. On the top row from left to right the panels show: 1) The distance between matched halo centres in  $h^{-1}\text{Mpc}$  2) The fractional difference in halo mass, as measured within a sphere of density 200 times the background density. 3 & 4) The fractional difference in the virial mass and radius of matched halos. On the lower row, from left to right, the panels show: 1) the fractional difference in the scale radius of matched halos. 2) The fractional difference in the velocity dispersion of matched halos. 3) The fractional difference in the magnitude of the velocities of the matched halos. 4) The difference in the direction of the velocities of matched halos in degrees. Hollow circles, filled circles, and stars represent the comparison for the 1296<sup>3</sup>, 1728<sup>3</sup>, and 2592<sup>3</sup> 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.



**Figure 8.** As for Fig. 7 but for halos matched to the high resolution simulation for each code. Note the change in range plotted in the bottom left panels.

not systematic effects being introduced by the nature of the halo matching algorithm.

The halo properties which are compared between matched catalogs are defined as follows. Halo separation is the distance between matched halos in units of  $h^{-1}\text{Mpc}$ , as discussed above this cannot be greater than  $0.25 h^{-1}\text{Mpc}$  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.  $r_{\text{vir}}$  is the radius of the halo which is similarly defined in terms of a spherical overdensity. In this case the overdensity is  $\Delta_c$  times the critical density of the universe.  $\Delta_c$  is the solution for a virialised cluster which is  $18\pi^2$  for a critical universe but can vary with redshift otherwise.  $M_{\text{vir}}$  is the mass contained within  $r_{\text{vir}}$  (Peebles 1980; Eke et al. 1996; Bryan & Norman 1998).  $r_s$  is the NFW scale radius of the halo, defined in terms of the density profile (Navarro et al. 1997).  $\sigma_v$  is the halo particle velocity dispersion in physical  $\text{km s}^{-1}$ .  $|v|$  is the speed of the halo in physical  $\text{km s}^{-1}$ . Finally  $\theta_v$  is the difference in angle of the velocity vectors of the matched halos in degrees. With the exception of halo separation,  $|v|$ , and  $\theta_v$ , 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 halos contain a greater number of particles and therefore it is to be expected that these halos 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  $100 h^{-1}\text{kpc}$ ; 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% in the halo mass and radius at medium and high resolution for all masses. At low resolution differences emerge for halo masses below  $10^{12} h^{-1}\text{M}_{\odot}$ .

$r_s$  shows a similar pattern of better agreement at higher halo masses. The differences at masses below  $10^{12} h^{-1}\text{M}_{\odot}$  are at the 10% level and do not show clear trends with code or resolution, except that SWIFT and GADGET show progressively lower  $\Delta r_s$  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 halos. There are large uncertainties on these measurements due to the limited number of high mass halos within each catalog.

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 halos 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% for all codes and resolutions at halo masses above  $10^{13} h^{-1}\text{M}_{\odot}$ . The same feature is seen in the virial radii where the difference is 0.2%. For lower mass halos 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 halos into low mass halos when compared to higher resolution simulations. The origin of this effect could be that the halo finder is more likely to split off halos 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 halos at all resolutions for all the codes. Below  $10^{12} h^{-1}\text{M}_{\odot}$ ,  $r_s$  becomes poorly matched as resolution is varied.

The velocity dispersion,  $\sigma_v$ , increases for lower mass halos 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 halos with mass greater than  $10^{13} h^{-1}\text{M}_{\odot}$  than lower mass halos.

Lower simulation resolution systematically biases high mass halos 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 halos show greater agreement in the matched halo parameters than low mass halos.

## 4.2 Halo Mass Function

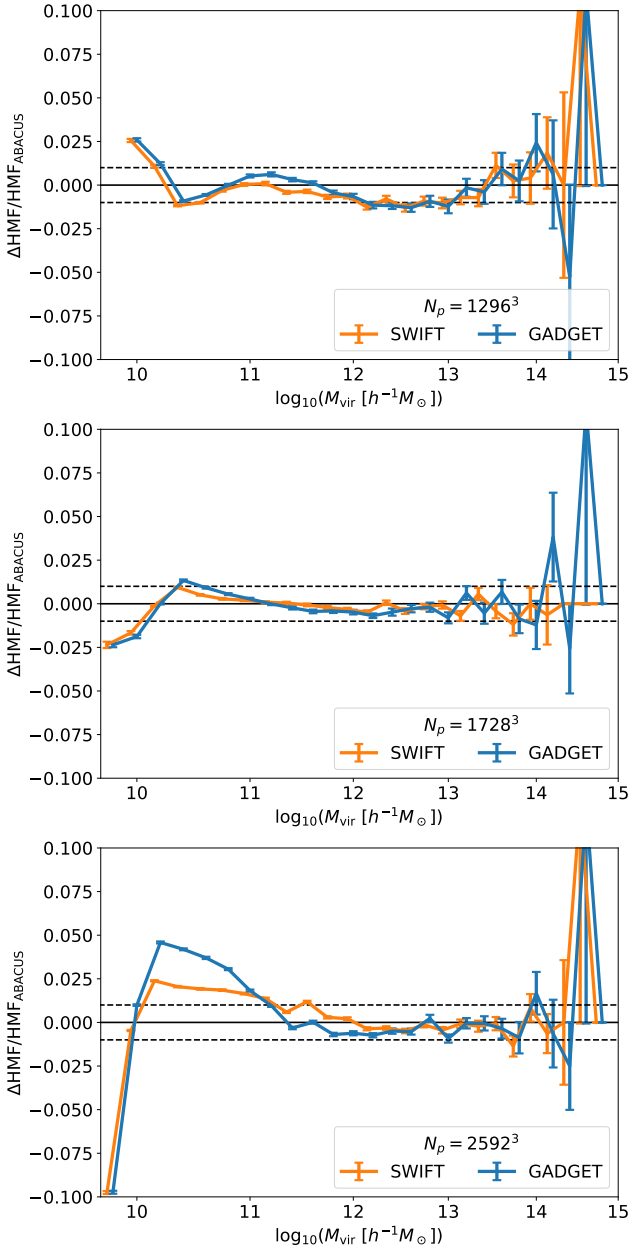
The halo mass function (HMF) describes the number density of dark matter halos in different mass intervals. Here we have defined the HMF as  $dn/d\log_{10}(M_{\text{vir}})$  where  $n$  is the number density of halos in units of  $\text{Mpc}^{-3}h^3$  and  $M_{\text{vir}}$  is in units of  $h^{-1}\text{M}_{\odot}$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

The HMF has a strong effect on the number density and clustering of galaxy catalogs created using the halo catalog, 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.

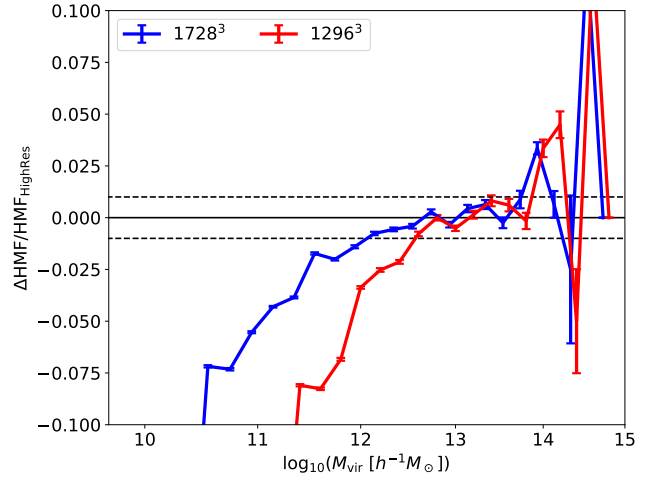
In addition we investigate the effect of simulation resolution on the HMF. Mass resolution has the greatest effect at the low mass end of the HMF where halos contain fewer particles. It is important to find the cutoff mass below which the HMF shows significant differences due to simulation resolution.

Figure 9 shows the HMFs measured from the different codes at fixed resolution at  $z = 1$ , relative to the ABACUS HMF. The errorbars 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% level between  $10^{12} h^{-1}\text{M}_{\odot}$  and  $10^{13.5} h^{-1} \text{M}_{\odot}$  for all resolutions. The HMFs do not differ by greater than 10% at masses up to  $10^{14.5} h^{-1} \text{M}_{\odot}$ , 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 catalogs from different resolution simulations have similar HMFs at for high mass halos. The medium resolution simulation HMF differs from the high resolution simulation HMF by more than 1% below  $10^{12} h^{-1} \text{M}_{\odot}$ . The low resolution HMF differs by a greater amount than the medium resolution HMF, with greater than 1% disagreement below  $10^{12.5} h^{-1} \text{M}_{\odot}$ .



**Figure 9.** The ratio of the halo mass functions 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 errorbars show an estimate for the noise while comparing two simulations which share the same phases. Differences between the codes are below 1% between  $10^{12} h^{-1} M_{\odot}$  and  $10^{13.5} h^{-1} M_{\odot}$  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, in the high resolution comparison the HMFs differ by up to 5% between  $10^{10} h^{-1} M_{\odot}$  and  $10^{11} h^{-1} M_{\odot}$  meanwhile in the medium and low resolutions the differences do not exceed 2.5% across the same mass range.



**Figure 10.** The ratio of the halo mass functions from the `ABACUS` simulations relative to the high resolution run at  $z = 1$ . The errorbars 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 halos in the lower resolution simulations relative to the high resolution simulation.

### 4.3 Halo Clustering: Correlation Function

The clustering of dark matter halos 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.

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

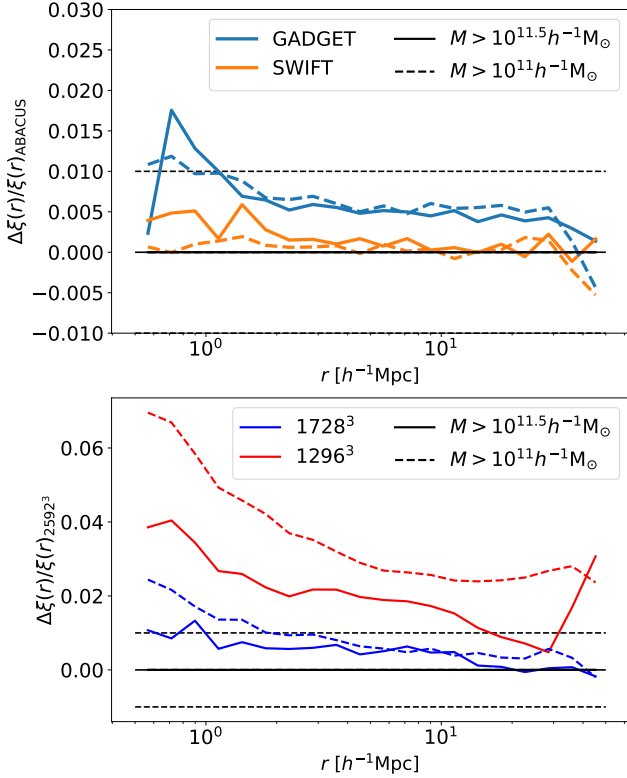
$$\xi(|\mathbf{r}|) = \langle \delta(\mathbf{x})\delta(\mathbf{x} - \mathbf{r}) \rangle \quad (9)$$

Fig. 11 shows the real-space two-point correlation functions of the halo catalogs from the different codes relative to `ABACUS`. The high resolution simulations were used with a mass cut of  $M > 10^{11.5} h^{-1} M_{\odot}$  and  $M > 10^{11} h^{-1} M_{\odot}$  (solid and dashed lines respectively). The correlation functions from the different simulation codes do not differ by greater than 1% for length scales above  $1 h^{-1} \text{Mpc}$ .

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

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

The multipoles of the redshift space correlation function,  $\xi_l(s)$



**Figure 11.** Upper Panel: Real-space 2-point correlation functions of the halos 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%. There are not significant differences which come from changing the halo mass limit. Lower Panel: Real-space two-point correlation functions of the halos 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 2-point correlation function.

are defined as

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

(Peebles 1980; Kaiser 1987). Here  $l$  is the multipole order,  $\theta$  is the angle between the line of sight and the halo separation vector,  $\mu = \cos(\theta)$ ,  $\xi(\sigma, \pi)$  is the 2d correlation function,  $\mathcal{L}_l(\mu)$  is the  $l^{\text{th}}$  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 sub-volumes. The redshift space correlation function multipoles were measured on  $N$  sub-samples created by removing a single sub-volume each time to

provide  $N$  clustering estimates. The standard error of the clustering can then be found to be:

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

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

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

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  $20 h^{-3} \text{Gpc}^3$  and for the year 1 DESI volume we use  $4 h^{-3} \text{Gpc}^3$  (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

$$\sigma_{\text{DESI}}(s) = \sqrt{\frac{V_{\text{sim}}}{V_{\text{DESI}}}} \sigma(s) \quad (13)$$

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

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 initial conditions, which removes sample variance. However our comparisons between different simulations with the same initial conditions still contain noise in the sense that the differences between these simulations contains a noise component that would change randomly if we changed the initial conditions 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 sub-samples

$$\sigma_{\text{NOISE}}(s) = \sqrt{\frac{N-1}{N} \sum_{i=0}^N (\Delta \xi^i(s) - \overline{\Delta \xi}(s))^2} \quad (14)$$

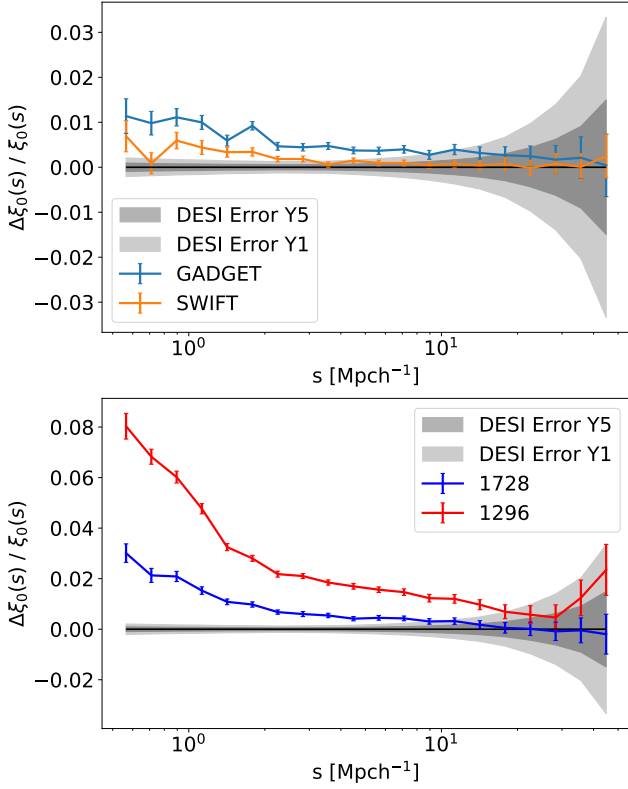
where  $\Delta \xi^i(s)$  is the difference in the clustering measurements from two simulations with the  $i^{\text{th}}$  sub-volume removed and  $\overline{\Delta \xi}(s)$  is the mean clustering difference. A detailed investigation of the jackknife method estimating the noise between two simulations sharing the same phases will be presented in Zhang et al. (2021).

$\sigma_{\text{NOISE}}(s)$  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  $10 \text{ Mpc} h^{-1}$ , with sub 0.1% 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



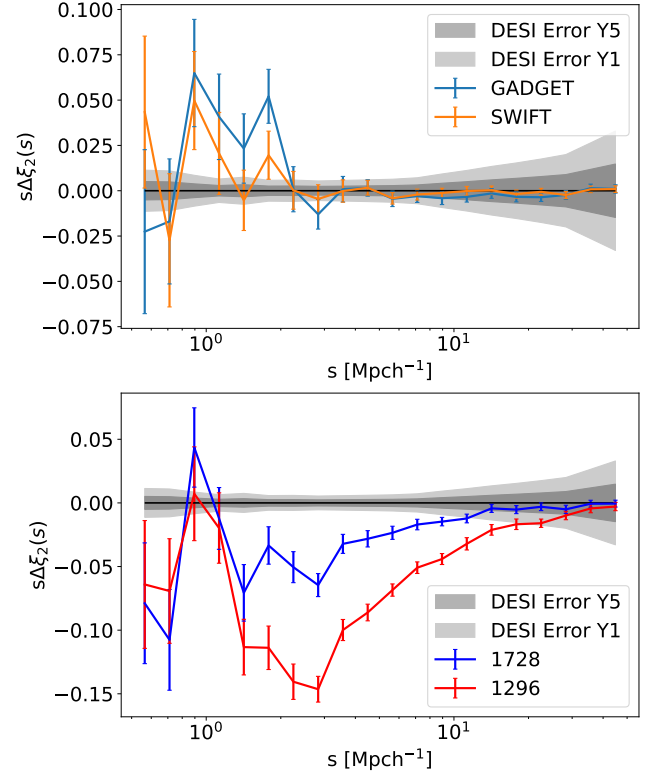


**Figure 12.** Upper panel: The redshift space clustering monopole comparison between the different codes at high resolution. *ABACUS* is used as the reference clustering. The shaded regions show an estimate for the DESI survey year 1 and year 5 statistical errors. The errorbars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of  $M > 10^{11.5} h^{-1} M_{\odot}$  was used. The systematic differences between the codes are within the expected year 5 DESI volume statistical errors for  $s > 20 h^{-1} \text{Mpc}$  and within the year 1 volume statistical errors for  $s > 10 h^{-1} \text{Mpc}$ . 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  $s > 20 h^{-1} \text{Mpc}$ . 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.

monopole and quadrupole difference between catalogs from different codes is smaller than the DESI statistical error above  $20 h^{-1} \text{Mpc}$ . *ABACUS* and *SWIFT* have remarkable agreements down to  $2 h^{-1} \text{Mpc}$ .

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  $s > 20 h^{-1} \text{Mpc}$  in the monopole and quadrupole.

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 halos in the regime where  $s > 20 h^{-1} \text{Mpc}$  for halos with mass larger than  $10^{11.5} h^{-1} M_{\odot}$ .



**Figure 13.** Upper panel: The redshift space clustering 2nd 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 errorbars show an estimate for the noise while comparing two simulations which share the same phases. A halo mass cut of  $M > 10^{11.5} h^{-1} M_{\odot}$  was used. The codes agree to within the DESI statistical errors for  $s > 10 h^{-1} \text{Mpc}$ , below which noise begins to dominate the observed differences. Lower panel: As above but showing the redshift space clustering 2nd 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  $s < 10 h^{-1} \text{Mpc}$ .

#### 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

$$P_l(k) = (2l + 1) \int_0^1 d\mu P(k, \mu) \mathcal{L}_l(\mu), \quad (15)$$

where  $l$  is the multipole order,  $\mu$  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,  $\mathcal{L}_l$  is the  $l^{\text{th}}$  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. \(2021\)](#).

In Fig. 14 the real-space power spectra are compared between codes. SWIFT, GADGET-2 and ABACUS agree at the 1% level for  $0.01 \leq k \leq 1 \text{ h Mpc}^{-1}$ . 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  $k = 0.2 \text{ h Mpc}^{-1}$ .

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

The differences between the codes and resolutions in the redshift space quadrupole are shown in Figure 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  $k < 0.05 \text{ h Mpc}^{-1}$ . 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 initial conditions generators and also  $N$ -body simulation codes.

### 5.1 IC Comparison

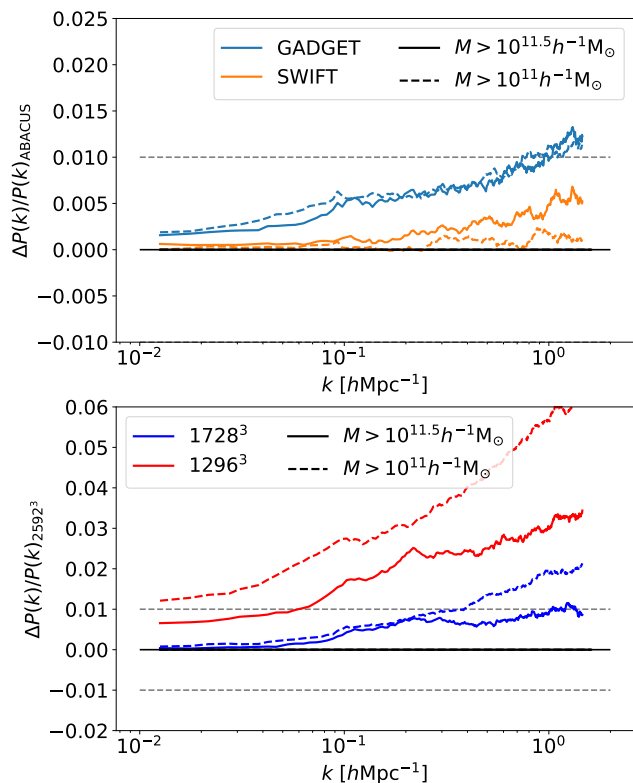
The initial conditions code comparison included the IC generators associated with the ABACUS, and FASTPM  $N$ -body simulation codes, along with the IC generator PANPHASIA/IC\_GEN. ICs were generated at three different resolutions for each code. Figure 1 shows that the power spectra of the phase matched ICs from the same code agreed to within 1% up to 20% 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 agree with the input linear power spectrum within 1% up to 50% of the Nyquist frequency. Although we chose PANPHASIA to setup 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 Figure 2. The codes were mostly consistent with one another to within the sample variance estimated by using ten realisations of the PANPHASIA 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, & FASTPM were run from identical PANPHASIA ICs from  $z = 199$  to  $z = 0$ . The matter power spectra were compared at  $z = 2$  and  $z = 1$ .

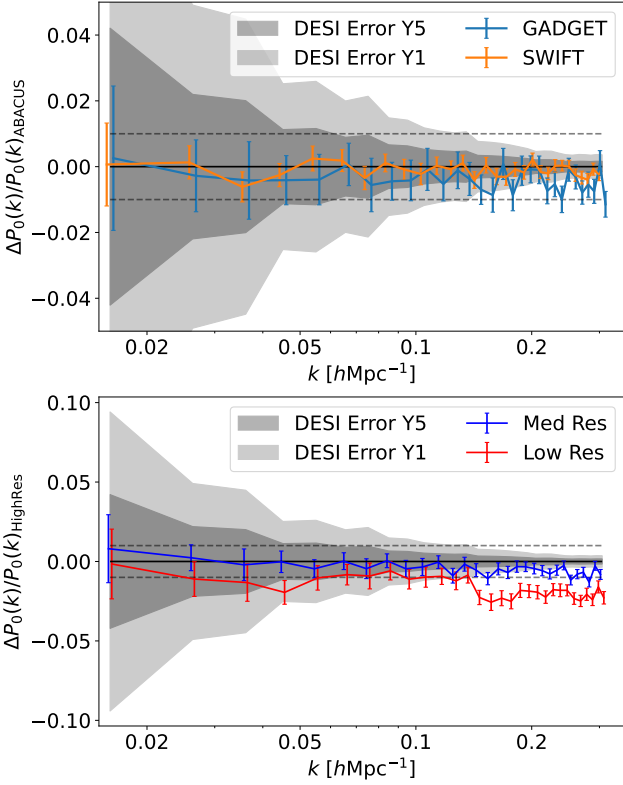


**Figure 14.** Upper panel: The ratio of the real-space power spectra of the halos from the high resolution simulations for different codes to ABACUS at  $z = 1$ . There is agreement between the codes to within the 1% level at both mass cuts for  $k < 1 \text{ h Mpc}^{-1}$ , with SWIFT and ABACUS showing agreement to within 0.1% for  $k < 0.1 \text{ h Mpc}^{-1}$ . Lower panel: The ratio of the real-space power spectra of the ABACUS halos compared to 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.

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 Figure 3, the power spectra agree to within 1% at  $k = 2 \text{ h Mpc}^{-1}$  and to within 4% at  $k = 10 \text{ h Mpc}^{-1}$  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% at large scales as seen in Figure 4. Figure 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% at  $k = 10 \text{ h Mpc}^{-1}$  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% level for  $k > 0.5 \text{ h Mpc}^{-1}$ . Similar conclusions are reached at the two explored redshifts.

The dark matter halos were found using ROCKSTAR. Halo properties were compared between matched halos for fixed code and variable resolution or fixed resolution and variable code and the results were shown in Figures 7 & 8. There was the greatest agreement between codes for high mass halos in most statistics. The difference



**Figure 15.** Upper panel: The ratio of the redshift space power spectrum monopoles of the halos 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 errorbars 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  $k < 0.1 h \text{ Mpc}^{-1}$ . 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  $k < 0.2 h \text{ Mpc}^{-1}$ . The low resolution simulation shows significant differences for  $k > 0.15 h \text{ Mpc}^{-1}$ .

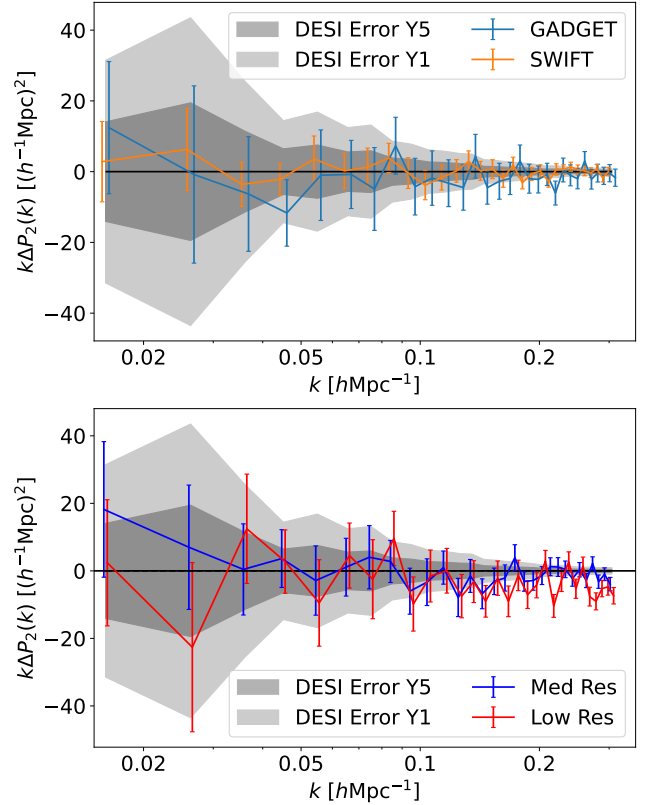
between matched halo properties was within 1% between all the codes for halo masses greater than  $10^{13} h^{-1} M_{\odot}$ .

Matching halo properties between different resolutions presented several systematic effects. Low mass halos in low resolution simulations had greater mass than the matched halos 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 halo mass functions from *SWIFT*, *GADGET-2*, and *ABACUS* agreed to within 1% between mass limits of  $10^{11.5} - 10^{14} h^{-1} M_{\odot}$  as seen in Figure 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  $10^{11.5} h^{-1} M_{\odot}$ . The minimum halo mass that hosts the DESI ELG 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  $10 h^{-1} \text{ Mpc}$  and within the DESI year 5 uncertainty at scales larger than  $20 h^{-1} \text{ Mpc}$  (Figs. 11-13). *ABACUS* and *SWIFT*



**Figure 16.** Upper panel: The difference of the redshift space power spectrum quadrupoles of the halos 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 errorbars 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  $k < 0.05 h \text{ Mpc}^{-1}$ , where the simulation noise is comparable in magnitude to the DESI errors, the code and resolution comparisons appear to be consistent with one another.

have remarkable agreements down to  $2 h^{-1} \text{ Mpc}$ . Comparing the different resolutions, the medium resolution run agrees with the high resolution one at scales larger than  $20 h^{-1} \text{ Mpc}$ . 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.  $k < 0.3 h \text{ Mpc}^{-1}$ . 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 HOD 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. 2021).

## 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 U.S. 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 Technologies 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 of Mexico; the Ministry of Economy of Spain, and by the DESI Member Institutions. The authors are honored 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. B. L. would like to acknowledge the support of the National Research Foundation of Korea (NRF-2019R111A1A01063740) 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](http://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 initial conditions files, simulation snapshots, and halo catalogs available at a Globus link<sup>11</sup>. 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, *Monthly Notices of the Royal Astronomical Society*, 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, *Monthly Notices of the Royal Astronomical Society*, 311, 793
- Bryan G. L., Norman M. L., 1998, *The Astrophysical Journal*, 495, 80
- Cole S., Hatton S., Weinberg D. H., Frenk C. S., 1998, *Monthly Notices of the Royal Astronomical Society*
- Conroy C., Wechsler R. H., 2009, *The Astrophysical Journal*, 696, 620
- Crocce M., Pueblas S., Scoccamarro R., 2006, *Monthly Notices of the Royal Astronomical Society*, 373, 369–381
- DESI 2016, The DESI Experiment Part I: Science, Targeting, and Survey Design ([arXiv:1611.00036](https://arxiv.org/abs/1611.00036))
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *The Astrophysical Journal*
- DeRose J., et al., 2019, *The Astrophysical Journal*, 875, 69
- DeRose J., et al., 2021, [arXiv:2105.13547](https://arxiv.org/abs/2105.13547)
- Dehnen W., 2014, *Computational Astrophysics and Cosmology*, 1
- Ding Z., et al., 2021, in preparation
- 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, *Monthly Notices of the Royal Astronomical Society*, 461, 4125
- Garrison L. H., Eisenstein D. J., Pinto P. A., 2019, *Monthly Notices of the Royal Astronomical Society*, 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, *The Astronomical Journal*, 156, 160
- Hernández-Aguayo C., et al., 2021, in preparation
- Jenkins A., 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 1859
- Jenkins A., 2013, *Monthly Notices of the Royal Astronomical Society*, 434, 2094
- Kaiser N., 1987, *MNRAS*, 227, 1
- Levi M., et al., 2013, The DESI Experiment, a whitepaper for Snowmass 2013 ([arXiv:1308.0847](https://arxiv.org/abs/1308.0847))
- Maksimova N. A., Garrison L. H., Eisenstein D. J., Hadzhiyska B., Bose S., Satterthwaite T. P., 2021, *MNRAS*,
- Marcos B., Baertschiger T., Joyce M., Gabrielli A., Labini F. S., 2006, *Physical Review D*, 73
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *The Astrophysical Journal*, 490, 493
- Norberg P., Baugh C. M., Gaztañaga E., Croton D. J., 2009, *Monthly Notices of the Royal Astronomical Society*, 396, 19
- Peacock J. A., Smith R. E., 2000, *Monthly Notices of the Royal Astronomical Society*, 318, 1144
- Peebles P. J. E., 1980, The large-scale structure of the universe. Princeton University Press
- Planck Collaboration et al., 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, *Monthly Notices of the Royal Astronomical Society*, 338, 14–34
- Rodríguez-Torres S. A., et al., 2016, *Mon. Not. Roy. Astron. Soc.*, 460, 1173
- Schaller M., Gonnet P., Chalk A. B. G., Draper P. W., 2016, *Proceedings of the Platform for Advanced Scientific Computing Conference*
- Schaller M., Gonnet P., Draper P. W., Chalk A. B. G., Bower R. G., Willis J.,

<sup>11</sup> [https://app.globus.org/file-manager?origin\\_id=53f3eea72-5215-11ec-a9cc-91e0e7641750&origin\\_path=%2F](https://app.globus.org/file-manager?origin_id=53f3eea72-5215-11ec-a9cc-91e0e7641750&origin_path=%2F)

- Hausammann L., 2018, SWIFT: SPH With Inter-dependent Fine-grained Tasking (ascl:1805.020)
- Schneider A., et al., 2016, *Journal of Cosmology and Astroparticle Physics*, 2016, 047
- Smith A., de Mattia A., Burtin E., Chuang C.-H., Zhao C., 2020, *MNRAS*, 500, 259
- Springel V., 2005, *Monthly Notices of the Royal Astronomical Society*, 364, 1105
- Springel V., et al., 2005, *Nature*, 435, 629
- Theuns T., Chalk A., Schaller M., Gonnet P., 2015, SWIFT: task-based hydrodynamics and gravity for cosmological simulations (arXiv:1508.00115)
- Wechsler R. H., Tinker J. L., 2018, *Annual Review of Astronomy and Astrophysics*, 56, 435
- Winther H. A., et al., 2015, *Monthly Notices of the Royal Astronomical Society*, 454, 4208–4234
- Wu C. F. J., 1986, *The Annals of Statistics*, 14, 1261
- Zel'Dovich Y. B., 1970, *A&A*, 500, 13
- Zhang H., et al., 2021, in preparation
- de Jong J. T. A., Kleijn G. A. V., Kuijken K. H., Valentijn E. A., 2012, *Experimental Astronomy*, 35, 25

<sup>20</sup>*Physics Department, University of Michigan Ann Arbor, MI 48109, USA*

## APPENDIX A: COMPARISON OF SUBHALO MASS FUNCTIONS

The comparison of subhalo mass functions between the simulations at fixed resolution at  $z = 1$  is presented in Figure A1. The differences between the subhalo mass functions are smaller at lower resolution, lying within 1%, 2%, and 5% in the low, medium, and high resolutions respectively between  $10^{11}$  and  $10^{12} h^{-1} M_{\odot}$ . SWIFT and GADGET-2 show good agreement within 1% for all resolutions between the mass ranges of  $10^{11}$  and  $10^{12} h^{-1} M_{\odot}$ . Above  $10^{13} h^{-1} M_{\odot}$  the subhalo mass functions become noisy. The subhalo mass function is likely to be more sensitive to differences in the softening schemes of the different simulations than the parent halo mass function.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.

## AFFILIATIONS

<sup>1</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, Durham, DH1 3LE, UK*

<sup>2</sup>*Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, USA*

<sup>3</sup>*Department of Physics and Astronomy, University of Utah, Salt Lake City, UT 84112, USA*

<sup>4</sup>*Instituto de Astronomía, Universidad Nacional Autónoma de México, Apdo. Postal 20-364, México*

<sup>5</sup>*Department of Physics, Manipur University, Canchipur, 795003, Manipur, India*

<sup>6</sup>*Center for Computational Astrophysics, Flatiron Institute Simons Foundation, 162 Fifth Ave. New York, NY 10010, USA*

<sup>7</sup>*Department of Physics and Astronomy, Sejong University, Seoul 05006, Korea*

<sup>8</sup>*Berkeley Center for Cosmological Physics, Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA*

<sup>9</sup>*Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str 1, D-85748 Garching, Germany*

<sup>10</sup>*Excellence Cluster ORIGINS, Boltzmannstrasse 2, D-85748 Garching, Germany*

<sup>11</sup>*Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*

<sup>12</sup>*Department of Physics, Kansas State University, Manhattan, KS 66506, USA*

<sup>13</sup>*Department of Astronomy, School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, People's Republic of China*

<sup>14</sup>*Center for Astrophysics | Harvard & Smithsonian, 60 Garden St., Cambridge, MA 02138, USA*

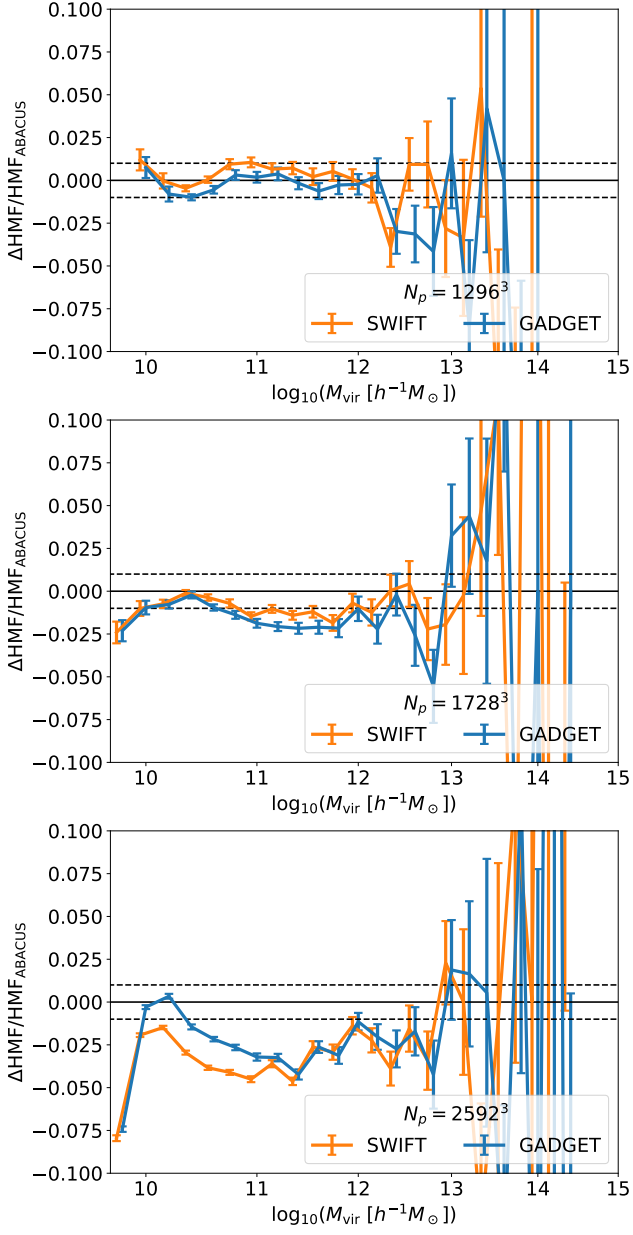
<sup>15</sup>*Centre for Extragalactic Astronomy, Department of Physics, Durham University, Durham, DH1 3LE, UK*

<sup>16</sup>*University College London, Department of Physics & Astronomy, Gower Street, London, WC1E 6BT*

<sup>17</sup>*Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA*

<sup>18</sup>*NSF's National Optical-Infrared Astronomy Research Laboratory, 950 N. Cherry Ave., Tucson, AZ 85719, USA*

<sup>19</sup>*Space Sciences Laboratory (SSL), UC Berkeley*



**Figure A1.** The ratio of the subhalo mass functions 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 errorbars show an estimate for the noise while comparing two simulations which share the same phases.