Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

High mass and halo resolution from fast low resolution simulations

Biwei Dai,[a] Yu Feng,[a] Uroš Seljak[a,b,c] and Sukhdeep Singh[a]

[a]Berkeley Center for Cosmological Physics and Department of Physics, University of California, Berkeley, CA 94720, U.S.A.
[b]Department of Astronomy, University of California, Berkeley, CA 94720, U.S.A.
[c]Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, U.S.A.

E-mail: biwei@berkeley.edu, yfeng1@berkeley.edu, useljak@berkeley.edu, sukhdeep1@berkeley.edu

Received August 16, 2019
Revised January 16, 2020
Accepted March 1, 2020
Published April 1, 2020

Abstract. Generating mocks for future sky surveys requires large volumes and high resolutions, which is computationally expensive even for fast simulations. In this work we try to develop numerical schemes to calibrate various halo and matter statistics in fast low resolution simulations compared to high resolution N-body and hydrodynamic simulations. For the halos, we improve the initial condition resolution and develop a halo finder "relaxed-FoF", where we allow different linking lengths for different halo mass and velocity dispersions. We show that our relaxed-FoF halo finder improves the common statistics, such as halo bias, halo mass function, halo auto power spectrum, cross correlation coefficient with the reference halo catalog, and halo-matter cross power spectrum. We also calibrate small-scale velocities of small halos to improve the power spectrum in redshift space. For the matter statistics, we incorporate the potential gradient descent (PGD) method into fast simulations to improve the matter distribution at nonlinear scales. By building a lightcone output, we show that the PGD method significantly improves the weak lensing convergence tomographic power spectrum. With these improvements FastPM is comparable to the high resolution full N-body simulation of the same mass resolution, with two orders of magnitude fewer time steps. These techniques can be used to improve the halo and matter statistics of FastPM simulations for mock catalogs of future surveys such as DESI and LSST.

Keywords: cosmological simulations, dark matter simulations, power spectrum, weak gravitational lensing

ArXiv ePrint: 1908.05276

 https://doi.org/10.1088/1475-7516/2020/04/002

JCAP04(2020)002

Contents

1 Introduction **1**

2 Halo statistics and clustering **2**
- 2.1 Relaxed-FoF 2
- 2.2 Halo statistics in real space 6
- 2.3 Velocity calibration and halo statistics in redshift space 10

3 Dark matter statistics **11**
- 3.1 PGD embedded in FastPM 11
- 3.2 Light-cone simulation 13
- 3.3 Weak lensing convergence 13

4 Conclusions **14**

A Halo power spectrum of higher mass thresholds **17**

1 Introduction

Numerical simulations of large scale structure formation are essential for extracting cosmological information from current and future sky surveys. N-body simulations with semi-analytic galaxy formation models have achieved great success in cosmological analysis [1–3], but they are also computationally expensive. Quasi N-body PM simulations with a small number of steps such as FastPM [4] and COLA [5] provide an alternative and fast way to model galaxy statistics. It has been shown that these fast simulations predict accurate halo statistics compared to full N-body simulations of the same resolution [4, 5]. However, to generate accurate mocks for future sky surveys such as DESI [6] and LSST [7], high mass resolution and large box volumes are needed, which makes the computational cost quite high even for fast simulations. For example, DESI aims at measuring the bright emission line galaxies up to $z = 1.7$, the analysis of which requires accurate modeling of $10^{11}h^{-1}M_\odot$ halos [8]. Considering that using halos with less than 200 particles could lead to large systematic errors [9], and that to cover the sky up to $z = 1.7$ the box should be around $3\,h^{-1}$ Gpc per side, we need at least 4 trillion dark matter particles in the simulation. This is computationally expensive in itself even with fast simulations like FastPM, not to mention that we may need lots of different realizations to measure the covariance matrices or to study the influence of cosmological parameters. Therefore, we need to find a model that reduces the computation cost while maintaining the accuracy.

Another difficulty in these quasi N-body simulations is the deficiency of their matter power on small scales due to insufficient force resolution. The potential gradient descent (PGD) model has been proposed to improve the modeling of matter distribution on nonlinear scales [10]. PGD was used as a post processing correction on the static snapshot. In this paper we incorporate PGD into FastPM at each time step, so that it can be used in generating time-continuous light-cone mocks for weak lensing analysis.

The goal of this paper is to produce reliable predictions for halo and dark matter statistics in low resolution FastPM simulations by training them on high resolution N-body simulations. The plan of the paper is as following. In section 2 we try to improve the identification of small halos by modifying the FoF halo finder and removing fake halos. The small-scale velocities of halos are calibrated to improve the modeling of redshift space distortion. For the matter field, we incorporate PGD into FastPM simulation in section 3. By building a light-cone output we show that the method can improve the weak lensing convergence field. Finally we conclude in section 4.

2 Halo statistics and clustering

In this section we examine and improve the halo statistics in FastPM simulation. We use IllustrisTNG [11] as our reference simulation. IllustrisTNG is a suite of cosmological hydrodynamic simulations with different box sizes and resolutions. We will mainly compare our results with TNG300-2-Dark, a dark-matter-only run in a $205\,h^{-1}$ Mpc periodic box and with 1250^3 particles. Since previous study shows that halo statistics have around 2% deviations for halos consisting of 200 particles [9], TNG300-2-Dark may not be accurate enough for the halo mass we consider (halos of 180 particles), so we also examine the TNG300-1-Dark simulation, which has a 8 times higher resolution. The agreement between TNG300-2-Dark and TNG300-1-Dark should give us an estimate of the accuracy of our reference simulation. Besides, we also show the results from TNG300-2 hydrodynamic simulation to study the baryonic effects on these halo statistics. All these simulations share the same initial linear density field.

To perform a direct comparison with the reference simulation, we run FastPM in the same $205\,h^{-1}$ Mpc periodic box with the same linear density field by matching the random seed and linear power spectrum. We generate the initial condition at $z = 9$ using 2LPT, and then evolve the field to redshift 0 with 40 steps distributed uniformly on the scale factor a. Unlike TNG300-2-Dark with 1250^3 dark matter particles, we run FastPM simulation with 8 times lower resolution, i.e., 625^3 particles. In this section we will mostly focus on $M > 10^{11}M_\odot = 6.7 \times 10^{10}h^{-1}M_\odot$ halos, corresponding to halos of more than 22 particles in FastPM. For most comparisons in this paper the halo catalogs are generated from abundance matching. In the following analysis, we also compare our results with TNG300-3-Dark, which has the same resolution but is a full N-body simulation. As we will see, with standard linking length $l = 0.2$, most of the comparisons with TNG300-3-Dark is quite good, consistent with previous study [4], but the agreement is not so satisfying when compared to the higher resolution reference simulation. This suggests that halos with less than a hundred particles cannot be modeled well even with a full N-body simulation, and the deficiency of FastPM at this mass range is mostly a resolution issue. This provides an additional motivation to our approach: by training on Illustris TNG300-2-Dark, which has a higher overall resolution (mass, time, and force), we can obtain results with FastPM that can exceed even Illustris TNG300-3-Dark despite its higher time and force resolution. We do so by modifying the standard Friends-of-Friends (FoF) algorithm to improve the situation for these small halos. All the halos in the Illustris TNG simulations are identified using FoF algorithm with linking length 0.2.

2.1 Relaxed-FoF

Large halos in FastPM can be modeled accurately, but small halos cannot be well resolved. For example, in figure 1 we show the same halo in the high resolution reference simulation, FastPM, and a full N-body simulation with the same resolution TNG300-3-Dark. We see

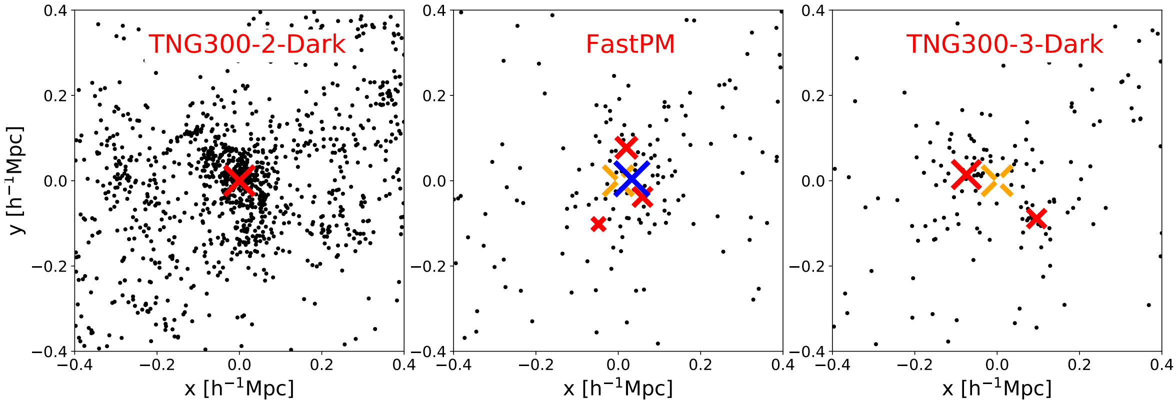

Figure 1. The projected map of the same halo in TNG300-2-Dark (left panel), FastPM (middle panel) and TNG300-3-Dark (right panel). The halo mass is around $1.6 \times 10^{11} h^{-1} M_\odot$, corresponding to 425 particles in TNG300-2-Dark simulation, and 53 particles in FastPM and TNG300-3-Dark. We perform the standard FoF algorithm with linking length 0.2 on each of them, and the halo centers of mass are represented as red crosses (the size of cross is proportional to the halo mass). The blue cross in the middle panel shows the center of mass of the halo identified with a larger linking length. In the middle and right panels we also show the true position (identified in TNG300-2-Dark) as orange dashed crosses.

that the lower mass resolution halo is more diffuse than the reference. The FoF algorithm with standard 0.2 linking length cannot link all the particles, and breaks the halo into 2 or 3 smaller halos. As a result, FastPM and TNG300-3-Dark tend to underestimate the halo mass at this mass range, and therefore underestimate the mass function, as we can see in figure 4. Because halos are broken into several small halos, they appear to be more clustered and produce a larger halo bias (figure 3).

We try to improve this situation by increasing the linking length l in the FoF halo finder as a function of halo mass. In the middle panel of figure 1, we see that with a larger linking length, we can successfully link all the particles and reproduce the correct halo mass and position. However, increasing the linking length for all the particles will bias the halo mass for large halos, since we know that the standard linking length $l = 0.2$ is already good for large halos at redshift 0 [4]. Therefore, we make the linking length a function of the halo particle number, with larger linking length for smaller halos. Since the linking length is not fixed, we call this method relaxed-FoF. As is shown in figure 4, the 0.2 linking length predict less massive halos at high redshifts, suggesting that the linking length should also be a function of redshift as well.

Another issue with these low resolution simulations is that in the high density regions, unbound clusters of nearby particles are linked by the FoF algorithm, therefore producing lots of fake halos. With larger linking lengths, we expect this issue to be more severe. We find that these fake halos are likely to have larger velocity dispersion. This is expected, since the particles that make up those fake halos are "accidental" close neighbors and are not gravitationally bound. Therefore, for each small halo we calculate the quantity $r = \frac{V_{\text{disp}}}{V_{\text{std,disp}}(M,z)}$, where V_{disp} is the velocity dispersion we measured from simulation, and $V_{\text{std,disp}}(M, z)$ is the expected velocity dispersion of a halo at this mass predicted by the common scaling relation [12].

$$V_{\text{std,disp}}(M, z) = V_0 \left(\frac{E(z) M}{10^{15} h^{-1} M_\odot} \right)^\alpha \tag{2.1}$$

procedure RelaxedFoF(x, N_p, l, r) ▷ x is the set of particles, N_p is the halo bin in ascending order, l is the corresponding linking length, and r is the velocity dispersion threshold.
    for $i \leftarrow 1, \mathrm{len}(N_p)$ **do**
        halo $\leftarrow$ FoF$(x, l[i])$
        for $j \leftarrow 1, N_{\mathrm{halo}}$ **do**
            if halo[j].Np $< Np[i]$ **and** halo[j].$V_{\mathrm{disp}} < r\ V_{\mathrm{std,disp}}$(halo[j].mass) **then**
                save halo[j] in halocat
                remove particles that form halo[j] from x
        remove particles that do not form halos from x
    $L \leftarrow l[\mathrm{len}(N_p)]$
    while x is not empty **do** ▷ keep reducing the linking length and save true halos & reject fake halos until no halos can be found
        $L \leftarrow 0.9L$
        halo $\leftarrow$ FoF(x, L)
        for $i \leftarrow 1, N_{\mathrm{halo}}$ **do**
            if halo[j].$V_{\mathrm{disp}} < r\ V_{\mathrm{std,disp}}$(halo[j].mass) **then**
                save halo[j] in halocat
                remove particles that form halo[j] from x
        remove particles that do not form halos from x

Algorithm 1. Relaxed-FoF algorithm.

where $V_0 \simeq 1100\,\mathrm{km\,s^{-1}}$, $E(z) = H(z)/H(0)$ is the dimensionless hubble parameter, and the slope α is around 0.3. If the quantity r is larger than a threshold r_0, we consider the halo as a fake one and reject it from the halo catalog. Since the fake halos we remove are mostly in the high density regions, we expect this procedure to reduce the bias of small halos.

We increase the linking length for small halos to better identify the halos in low resolution quasi-N-body simulations, and use velocity information to help remove the misidentified halos. Several previous papers have made similar attempts and achieved good results. For example, [13] generated PTHalos from a 2LPT field using a much larger linking length ($b = 0.38$), and then calibrated the halo mass to match the given halo mass function. Iteratively decreasing the linking length is commonly used in sub-halo finders (e.g., HFOF [14], ROCKSTAR [15]). Particle velocities have been incorporated in many modified FOF halo finders and phase space finders (e.g., 6DFOF [16], HSF [17], ROCKSTAR [15]), and many algorithms remove particles that are not dynamically bound to the halos. Considering that the small-scale velocities are not very accurate in our simulation, here we do not try to find halos in the phase space, and only use the dynamical information to decide whether a halo is real or not. We also tried removing unbounded particles from the halos, but we did not find improvements in our case.

We divide the halos into several bins $N_{p,i}$ ($N_{p,i}$ is the maximum halo particle number of bin i), and for each bin we have the corresponding linking length $l(N_{p,i}, z)$. We first run the FoF halo finder on all the particles in the snapshot with linking length $l(N_{p,1}, z)$ for the smallest halo bin. Then we select all the halos that are larger than the halo particle number

$N_{p,1}$ and the halos that are rejected by the velocity dispersion criterion, and rerun the FoF halo finder on the particles that form these halos with the linking length $l(N_{p,2}, z)$ of the next bin. We repeat this procedure until we finish the largest halo bin. For the rest of the particles that form the fake halos, we keep running the FoF halo finder and reducing the linking length, with fake halos rejected at each iteration, until there are no particles left. The function $l(N_{p,i}, z)$ and $r_0(z)$ are simple functions we choose to produce correct halo mass function and halo bias:

$$N_{p,i} = \{20, 40, 80, 160, 320, \inf\} \tag{2.2}$$

$$l(N_{p,1}, z) = l_1 - \frac{A_1}{1+z} \tag{2.3}$$

$$l(N_{p,6}, z) = \max\left(l_6 - \frac{A_2}{1+z},\ 0.2\right) \tag{2.4}$$

$$l(N_{p,i}, z) = \frac{(6-i)N_{p,1} + (i-1)N_{p,6}}{5} \tag{2.5}$$

$$r_0(z) = B_1 - B_2 \log(1+z) \tag{2.6}$$

where l_1, l_6, A_1, A_2, B_1 and B_2 are free parameters. In our setup we find $l_1 = 0.25$, $l_6 = 0.235$, $A_1 = 0.012$, $A_2 = 0.06$, $B_1 = 4.28$ and $B_2 = 2.17$ give us good halo statistics for $0 \le z \le 2$.

Even though relaxed-FoF calls a standard FoF algorithm more than 6 times, it does not take 6 times longer, because after each iteration the number of remaining particles quickly decreases. For example, only about 50% particles are left after the first iteration. In practice, we find that relaxed-FoF normally takes around twice as much time as standard FoF.

In addition to improving the halo finder algorithm, we also find that the small scale power in the initial condition is crucial for the identification of small halos. We find it necessary to generate the linear density map with a mesh that is twice finer than the particle grid, which helps to improve the various halo statistics (figure 2 to figure 6). We tried further increasing the resolution of the initial condition, but the halo statistics do not improve. Note that in this study the force resolution is also twice the resolution of particles, so we can use the same force mesh to generate initial condition, and increasing the resolution of IC does not require more memory than standard FastPM.

Before we examine any halo statistics in the next subsection, we first take a look at how well each individual halo can be reproduced. If two halos from two simulations are within $0.4\,h^{-1}$ Mpc and if their mass are within a factor of 2, then we say they are the same halo, and each halo cannot be matched with more than one halo. In figure 2 we show the ratio of missed halos as a function of halo mass for different redshifts. We define missed halos as the halos that cannot find a counterpart in the other simulation. We see that as we go to smaller halos, the ratio of missed halos increases and reaches around 18% (25% if the linear density map has the same resolution as particle grid) at $10^{11}h^{-1}M_\odot$ halos for FastPM with constant 0.2 linking length. After switching to our relaxed-FoF, the ratio of missed halos decreases at all redshifts and all halo masses. The improvement is larger at higher redshift. In particular, the ratio of missed halos is reduced to about 6% for $10^{11}h^{-1}M_\odot$ halo at redshift 2, comparable to the full N-body simulation TNG300-3-Dark.

In figure 2 we also show the ratio of missed halos for higher resolution N-body simulation TNG300-1-Dark and hydro simulation TNG300-2. Here the ratio is not zero even for large halos, due to the bridging effect. If two nearby halos are linked together in one simulation,

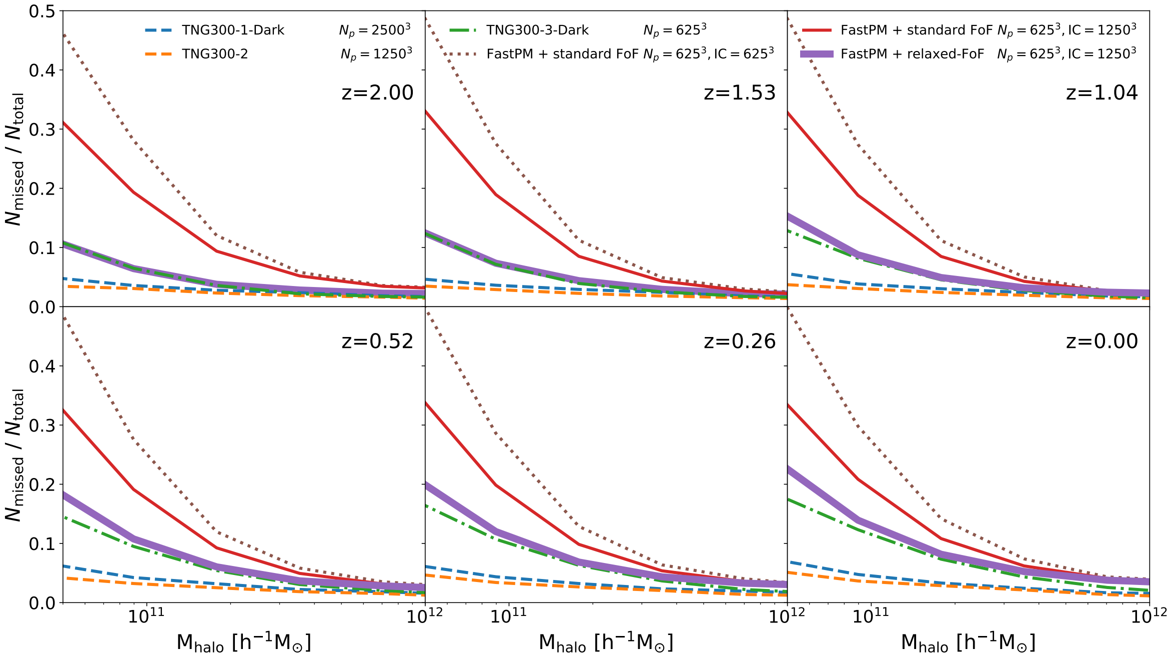

Figure 2. The ratio of missed halos as a function of halo particle number in different simulations and various redshifts. Here we choose TNG300-2-Dark as the reference (same for figure 3 to figure 8). The nonzero ratio of missed halo in the massive end is mostly because of the bridging effect.

while they are identified as two separate halos in another simulation, they will not be matched using our algorithm and therefore produce a nonzero fraction of missed halos.

2.2 Halo statistics in real space

We first examine the halo bias defined with the halo-matter cross correlation

$$b = \lim_{k \to 0} \frac{P_{hm}(k)}{P_{mm}(k)}. \tag{2.7}$$

We present the halo bias results in figure 3. We see that the bias given by different simulations fluctuate a lot even for the largest halos (lowest abundance). This is because the halo mass is scattered in different simulations so the same abundance does not guarantee the same halo catalog. However, comparing the three N-body simulations of different resolutions, we can see a tendency that higher resolution simulation shows a lower halo bias, especially for small halos. Similarly, FastPM also gives a very high bias for small halos, mostly due to its low resolution, but with our relaxed-FoF halo finder the halo bias is brought down to the normal level.

With larger linking length, the small halos can be better identified, so the improvement of the halo mass function at the low mass end is expected (shown in figure 4). With the same linking length 0.2, FastPM shows a large discrepancy with full N-body simulations of all resolutions, suggesting that this deficiency in halo mass function is not due to the resolution effect, but a failure of the FoF with standard linking length to resolve small halos in FastPM, which relaxed-FOF corrects for. Another interesting feature is that the baryonic feedback seems to reduce the mass function by 10% to 20%, but this comparison is based on FoF mass and it is unclear if it is a meaningful comparison against hydrodynamic simulations.

JCAP04(2020)002

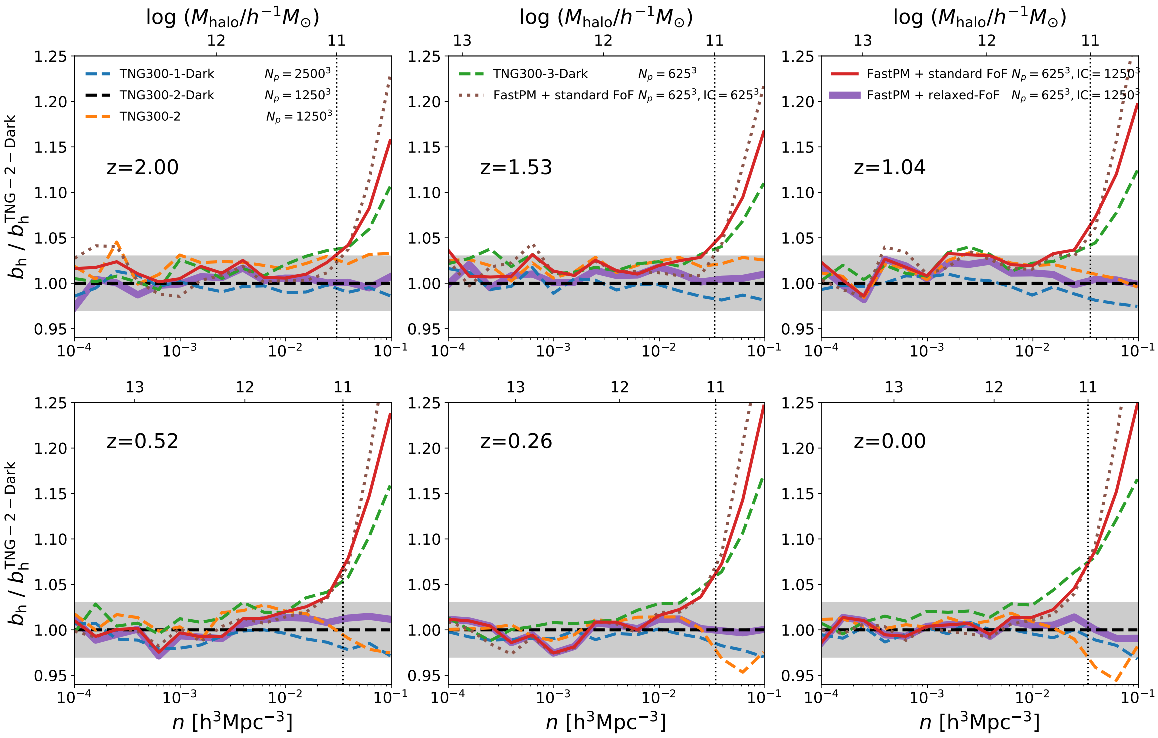

Figure 3. The ratio of halo bias as a function of abundance measured in different simulations and various redshifts. The halos are selected by abundance matching, and the x axis shows the halo abundance. Larger abundance means smaller halos. The dotted vertical line shows the abundance of a $10^{11}h^{-1}M_\odot$ halo, corresponding to 33 particles. The shaded region represents 3% deviation. The power spectrum is calculated using Nbodykit throughout the paper [18].

Next we select halo catalogs from different simulations with abundance matching, and examine their auto power spectrum (figure 5), halo-matter cross power spectrum (figure 6), and the cross correlation coefficient with the reference simulation TNG300-2-Dark (figure 7). The halo catalogs correspond to $M \geq 10^{11}M_\odot = 6.8\times10^{10}h^{-1}M_\odot$ halos (22 particles in FastPM). As mentioned above, the same abundance does not guarantee same halos, so we expect to see a little scatter across different simulations. As long as the deviation of FastPM is comparable to the scatter of TNG-1-Dark or TNG-2, we can say the predictions of FastPM are equivalent to those of more expensive high resolution full N-body simulations.

We see that with standard $l = 0.2$ linking length, the halo auto power spectrum and the halo-matter cross power spectrum of FastPM are very similar (slightly worse) to those of the full N-body simulation with the same resolution. After changing to relaxed-FoF, the halo auto power spectrum and halo-matter cross power spectrum improve on all scales and all redshifts. In particular, their deviations from TNG300-2-Dark are consistent with the scatter induced by abundance matching at high redshift ($z \geq 1$), while at low redshift ($z \leq 0.5$) on small scales FastPM is overpredicting power. Note that at low redshift, even though the auto power spectrum of FastPM is not consistent with our reference N-body simulation on small scales, its slope is actually quite similar to the prediction of TNG300-2 hydrodynamical simulation. The cross correlation coefficient is consistent with the ratio of missed halos (figure 2), that after improving the halo finder the halo catalog from FastPM is similar to a full N-body simulation TNG300-3-Dark.

JCAP04(2020)002

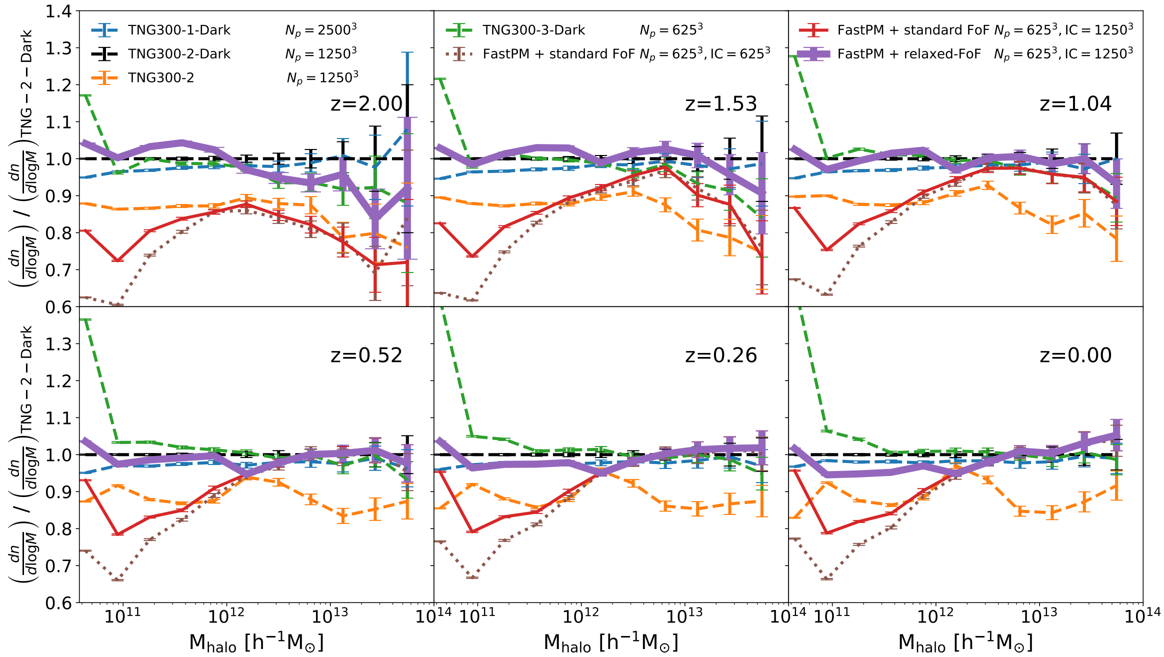

Figure 4. The ratio of halo mass function from different simulations and various redshifts. The halo mass here is defined as the FoF mass. For the hydro simulation the FoF algorithm is run on the dark matter particles, and baryon particles are attached to the same groups as their nearest dark matter particle. Poisson noise errors are shown.

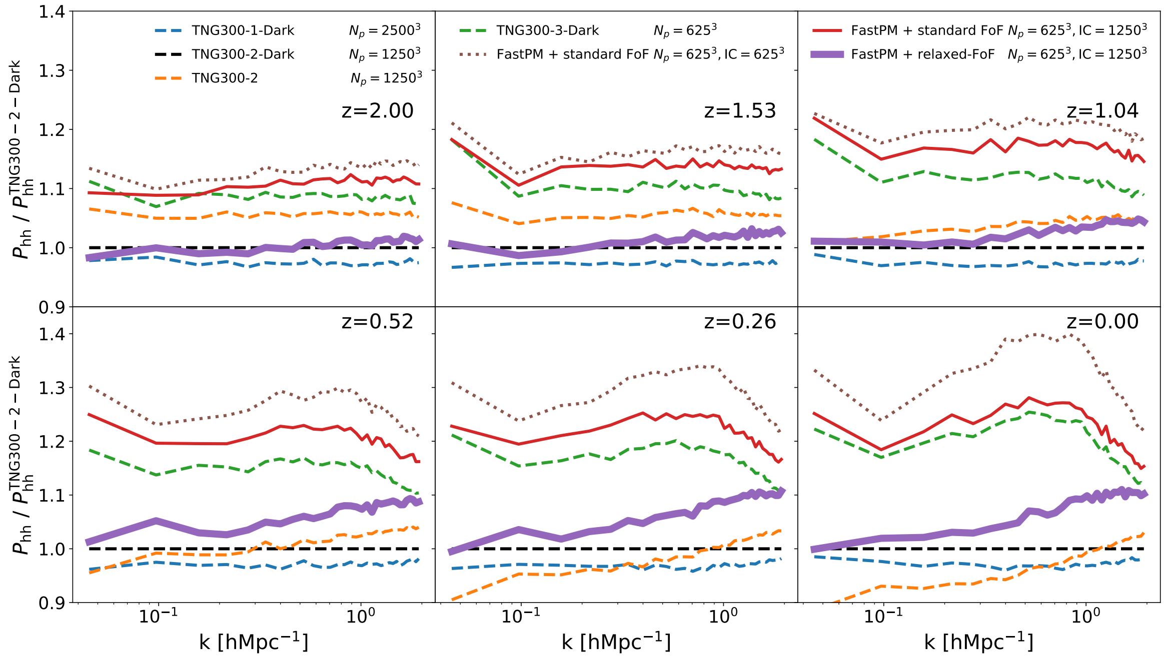

Figure 5. The ratio of halo auto power spectrum from different simulations and various redshifts. Similar to figure 2, TNG300-2-Dark is chosen as our reference. The halos are selected using abundance matching, corresponding to $M \geq 10^{11} M_\odot = 6.8 \times 10^{10} h^{-1} M_\odot$ halos (22 particles in FastPM).

JCAP04(2020)002

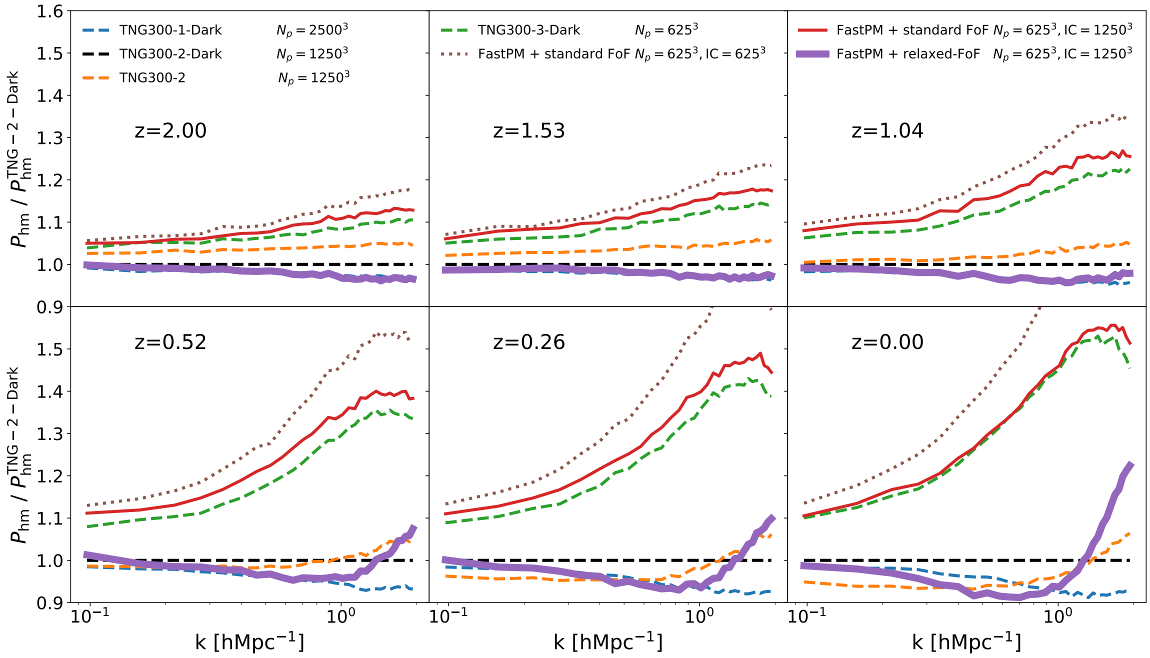

Figure 6. The ratio of halo-matter cross power spectrum from different simulations and various redshifts.

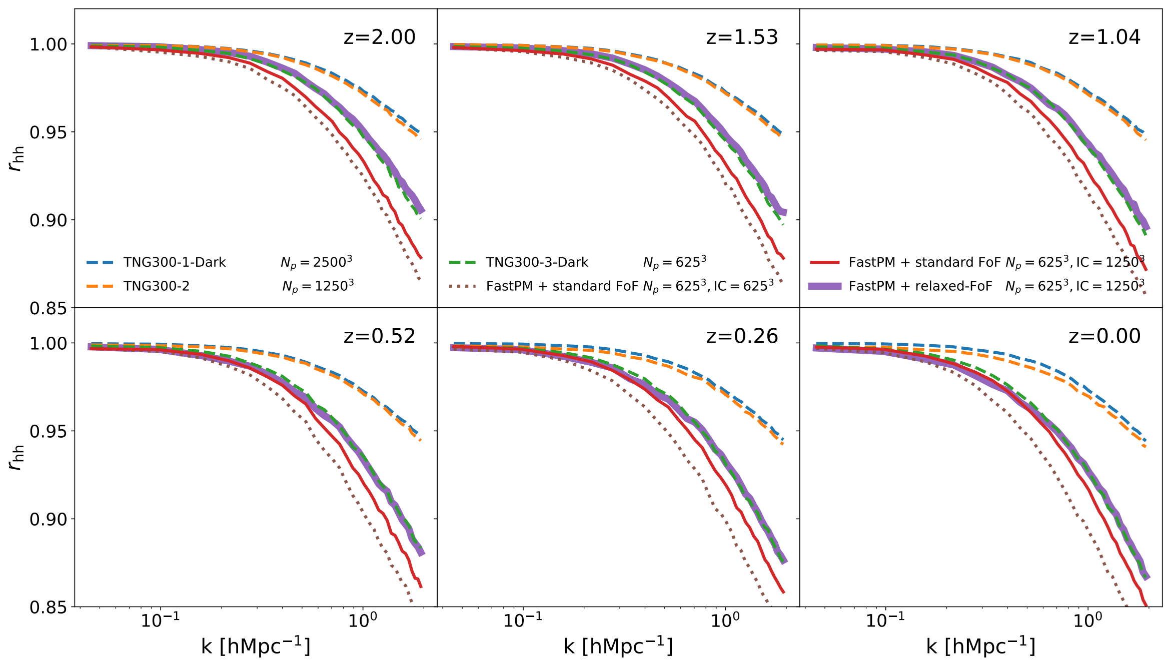

Figure 7. The cross correlation coefficient of the reference halo with halos from other simulations in different redshifts.

Here in this section we only show the results of $M \geq 10^{11} M_\odot = 6.8 \times 10^{10} h^{-1} M_\odot$ halos. The auto power spectrum of larger halos is shown in appendix A (figure 12 and figure 13).

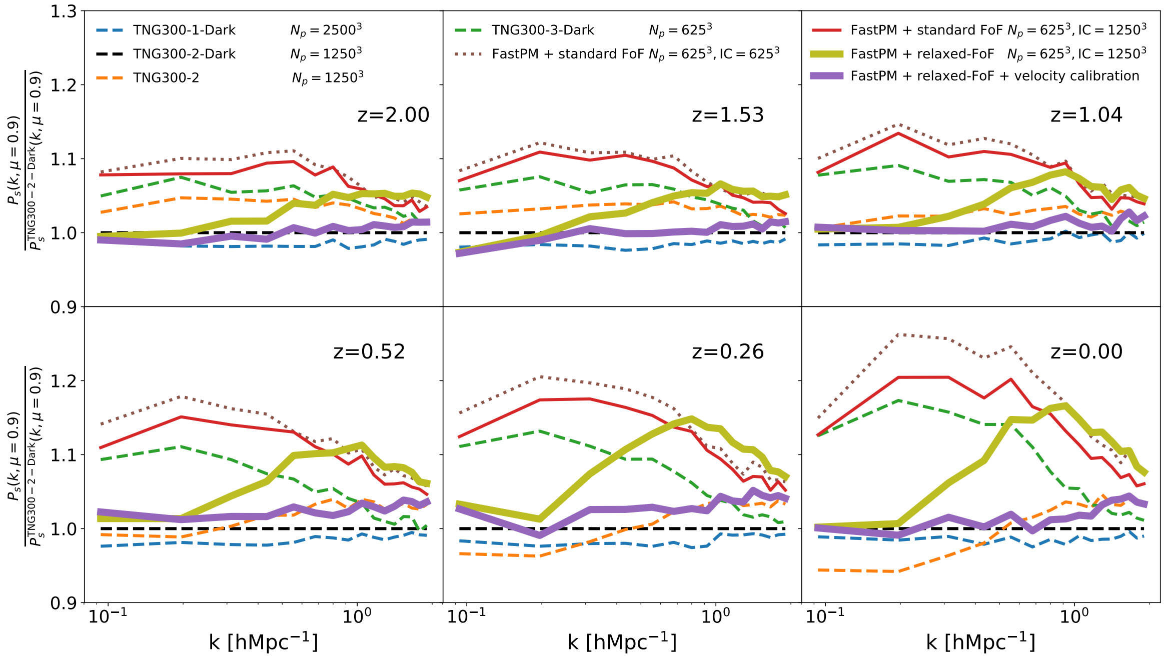

Figure 8. The ratio of halo power spectrum in redshift space from different simulations and various redshifts. Here we only show the power spectrum of k parallel to the line of sight, i.e., $\mu = 0.9$, where $\mu = k_{\parallel}/k$. The k mode perpendicular to the line of sight is not affected by RSD, and therefore the RSD halo power spectrum with $\mu = 0$ is similar to the halo power spectrum in real space presented in figure 5.

2.3 Velocity calibration and halo statistics in redshift space

In figure 8 we show the halo auto power spectrum in redshift space. On large scales, similar to the situation in real space, relaxed-FoF improves the power. On small scales, however, the velocities of small halos cannot be accurately modeled by low resolution FastPM. Even with relaxed-FoF, FastPM predicts too much power on small scales in redshift space. We try to calibrate the velocities by adding an irrotational velocity term to the particles inside small halos:

$$\mathbf{u}_{\mathrm{calib}} = \mathbf{u}_{\mathrm{COM}} - \left\langle \frac{1}{a} \nabla V \right\rangle \exp(-N/N_c) \tag{2.8}$$

where $\mathbf{u}_{\mathrm{COM}}$ is the measured halo center of mass velocity, N is the number of particles in the halo, and $N_c = 160$ is the halo size which we believe can be modeled well by FastPM (shown in appendix A). We do not intend to modify the velocities of the large halos, so we put a factor $\exp(-N/N_c)$ in the equation. Since we assume the new velocity component is irrotational, it can be written as the gradient of velocity potential V, and we average over all the particles in the halo to give a center-of-mass velocity correction. We try to learn the velocity potential V from the matter gravitational potential Φ, by assuming

$$V(a, k) = T(a, k)\Phi(a, k) \tag{2.9}$$

where a is the scale factor, and the transfer function $T(a, k)$ should go to zero when k goes to zero to prevent any modification of large-scale velocities. We further assume

$$T(a, k) = (C_1 a - C_2) \exp\left(-k_c^2/k^2\right) \exp\left(-k^2/k_s^2\right) \tag{2.10}$$

JCAP04(2020)002

where k_c, C_1 and C_2 are free parameters. We find $k_c = 1\mathrm{h}^{-1}\,\mathrm{Mpc}$, $C_1 = 0.23$ and $C_2 = 0.03$ give us good small-scale halo power in our setup. We introduce the factor $\exp(-k^2/k_s^2)$ to reduce the numerical effect induced by the mesh resolutions, and k_s is fixed to $5\mathrm{h}^{-1}\,\mathrm{Mpc}$. The auto power in redshift space of halos larger than $10^{11}M_\odot = 6.8\times10^{10}h^{-1}M_\odot$ is shown in figure 8. The results of other mass bins are shown in appendix A (figure 14 and 15). Noting the similarity between equation (2.8), (2.9), (2.10) and equation (3.1), (3.2), our velocity calibration model can be seen as applying Potential Gradient Descent (PGD) model to velocities, where both methods model correction vectors as the gradient of modified gravitational potential.

3 Dark matter statistics

In this section we focus on improving the matter distribution on small scales. We first incorporate PGD model into every step of FastPM, and show that the redshift evolution of the PGD parameters can be parametrized by simple analytical functions. Then we build a light-cone simulation with the output of FastPM, and show that the PGD model can improve the weak lensing convergence power spectrum.

3.1 PGD embedded in FastPM

The basic idea of the potential gradient descent (PGD) model is to add an additional displacement on the output position of particles to mimic the missing sub-grid physics in simulations. The additional displacement is modeled by the gradient of a modified gravitational potential, given by

$$\begin{aligned}\mathbf{S} &= (\alpha/H_0^2)\ \nabla(\hat{\mathbf{O}}_\mathbf{l}\hat{\mathbf{O}}_\mathbf{s}\phi)\\ &= (4\pi G\bar{\rho}\alpha/H_0^2)\ \nabla(\hat{\mathbf{O}}_\mathbf{l}\hat{\mathbf{O}}_\mathbf{s}\nabla^{-2}\delta)\end{aligned} \tag{3.1}$$

where ϕ is the gravitational potential field, δ is the matter overdensity, $\bar{\rho}$ is the averaged matter density, α is a free parameter, $\hat{\mathbf{O}}_\mathbf{l}$ and $\hat{\mathbf{O}}_\mathbf{s}$ are a high pass filter and a low pass filter, respectively

$$\hat{\mathbf{O}}_\mathbf{l}(k) = \exp\left(-\frac{k_l^2}{k^2}\right), \tag{3.2}$$

$$\hat{\mathbf{O}}_\mathbf{s}(k) = \exp\left(-\frac{k^4}{k_s^4}\right). \tag{3.3}$$

Here k_l and k_s are also free parameters.

[10] showed that the PGD model improves the halo profiles and the small scale power spectrum. The PGD model can be treated as a single post processing correction on the output snapshot, but it would be hard to do the correction on a lightcone output in this way, since the correction parameters are functions of redshift, and the redshift is not fixed in a lightcone output. Here we try to solve the problem by incorporating PGD correction into FastPM. We perform a PGD correction after each FastPM step, and then feed the corrected particle position into the next time step. Because the PGD is coupled into the simulation, both of the static snapshots and the lightcone output are consistently corrected.

JCAP04(2020)002

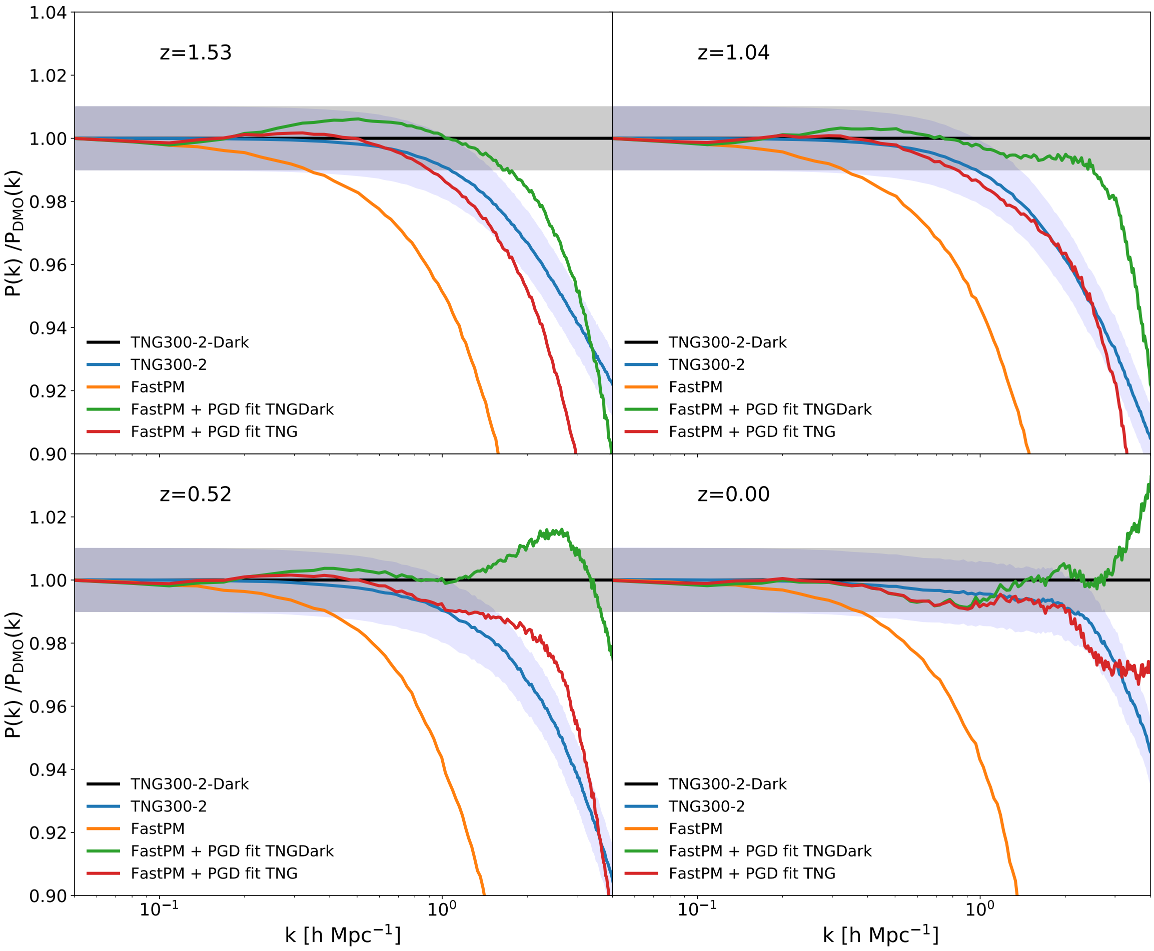

Figure 9. The matter power spectrum of FastPM simulation, before and after the calibration, and the reference simulation (TNG300 and TNG300-Dark) in different redshifts. Here we are trying to match the power spectrum of both TNG300-Dark and TNG300 (to account for the baryonic effects). The shadow region shows the 1% deviation. The resolution of FastPM is 125 times lower than TNG300-2-Dark.

JCAP04(2020)002

The parameters α and k_l are functions of redshift, and we model their redshift dependence by

$$\log\left(\frac{\alpha}{\alpha_0}\right) = Aa^2 - Ba, \tag{3.4}$$

$$k_l = k_{l,0}a^{\gamma}, \tag{3.5}$$

where α_0, A, B, $k_{l,0}$ and γ are free parameters. $k_s = k_{s,0}$ is another free parameter and is fixed for all redshifts. These parameters are fitted by matching the matter power spectrum at all redshifts simultaneously. In figure 9 we show the matter power spectrum of original FastPM, and FastPM after the correction. Unlike [10] where they are comparing the same resolution simulation, here the mass resolution of FastPM is 125 times lower than the TNG300-2-Dark, yet we show that we can match the power spectrum quite well. The cross correlation coefficient also improves on all redshifts, e.g., it improves approximately from 0.5 to 0.6 at the scale of $k = 10\,h^{-1}\,$Mpc.

3.2 Light-cone simulation

To test how the PGD model improves the weak lensing map, We build a light-cone output from the FastPM simulation. The FastPM simulation is run in a $3200\,h^{-1}$ Mpc periodic box with 1536^3 dark matter particles. The simulation starts at redshift 9, with time steps separated by constant spacing in the scale factor. The FastPM without PGD correction has 40 time steps, while after implementing the PGD correction we reduce the time steps to 20, since it has a comparable computation cost as a 40 step FastPM. The positions and velocities of the particles located between the steps are interpolated from nearest steps and are saved as the particle positions intersect the observer's light-cone. An optional FoF halo finder can be ran on the fly as the light-cone is generated, which uses padding to handle the continuity between light-cone slices. Given that the volume of the simulation box can be smaller than the light-cone, the simulation box is tiled (duplicated) as necessary to cover the required volume of the lightcone. The light-cone module of FastPM allows configurations on the position of the observer, the field of view angle (determines the sky-area), the direction of sightlines, the replication (tiling) matrix, the list of culling octants, and the redshift range of interest.

In this work we assume the observer sits at the origin of the simulation box, and integrate the sightlines up to $z = 2.2$. Note that the comoving distance to redshift 2.2 is around $3.8\,h^{-1}$ Gpc, and the box is replicated along all directions to include the extra $600\,h^{-1}$ Mpc.

3.3 Weak lensing convergence

Under the Born approximation, we estimate the weak lensing convergence map produced by the source galaxies between redshift $z_{\min}$ and $z_{\max}$ as (see e.g. [19, 20])

$$\kappa(\theta) = \frac{3H_0^2\Omega_m}{2c^2}\int_{z_{\min}}^{z_{\max}} p(z_s)dz_s \int_0^{\chi_s(z_s)} d\chi \frac{(\chi_s(z_s)-\chi)\chi}{\chi_s(z_s)a}\delta(\chi,\theta) \tag{3.6}$$

where θ is a 2D angular vector, δ is the matter overdensity at radial comoving distance χ and angular position θ, $\chi_s(z_s)$ is the comoving distance to redshift z_s, and $p(z_s)$ is the normalized redshift distribution of source galaxies between redshift $z_{\min}$ and $z_{\max}$. Weak lensing maps are generated in the post processing after the simulation has ended and the particle lightcone has been saved. In the post processing step, the lightcone particles are read in, integrated along the line of sight using the weights based on lensing kernel and then a pixelized map is generated with NGP (nearest grid point) window function using HEALPY [21], the python version of healpix [22]. While there is one I/O overhead due to lensing maps being generated in this manner, the post processing provides flexibility to generate multiple lensing maps for different lensing source configurations and saving the lightcone also allows one to generate lightcones for different probes in general (the line of sight integration kernel can be different from lensing kernel). Furthermore, for the case of cross correlations, it is also possible to generate the maps integrating over a narrow lens redshift range rather than over a complete redshift range from sources to observers, if necessary to reduce the I/O load.

In this work we assume a source galaxy redshift distribution of a LSST-like survey (second panel of figure 10). We divide the source into 3 tomographic bins: $z \in [0, 0.7]$, $[0.7, 1.4]$ and $[1.4, 2.1]$, and generate the convergence maps produced by these 3 source bins separately. In figure 10 we show the all-sky convergence map as well as the zoomed-in maps of the last tomographic bin. We see that the PGD correction makes the peaks more evident. Thus we expect that the correction to help with non-Gaussian statistics such as

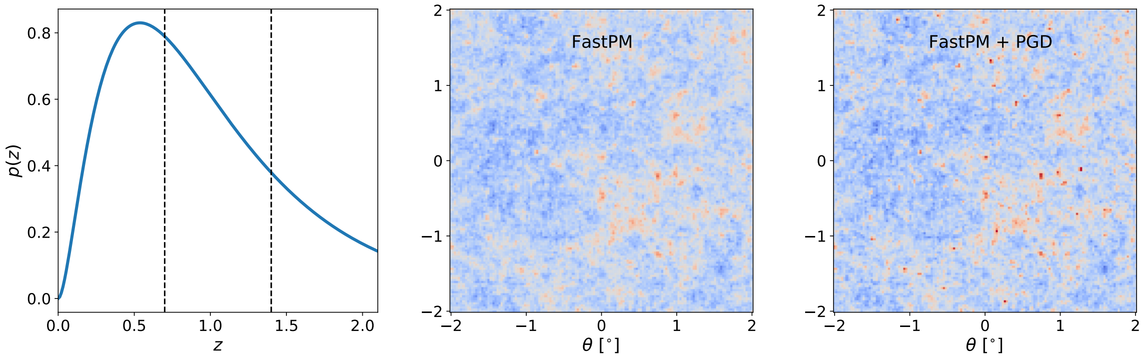

Figure 10. The redshift distribution of source galaxies of a LSST-like survey (left panel), and the zoomed-in convergence field of FastPM (middle panel) and FastPM with PGD correction (right panel).

peak statistics. Below we will examine the auto power spectrum and cross power spectrum of these convergence maps.

Under the Limber approximation, the angular power spectrum of the weak lensing convergence can be written as

$$\begin{aligned} C^{\kappa}(l) = \left(\frac{3H_0^2\Omega_m}{2c^2}\right)^2 \int_{z_{\min1}}^{z_{\max1}} p_1(z_{s,1})dz_{s,1} \int_{z_{\min2}}^{z_{\max2}} p_2(z_{s,2})dz_{s,2} \\ \int_0^{\chi_s(\min(z_{s,1},z_{s,2}))} d\chi \left(\frac{\chi_s(z_{s,1})-\chi}{\chi_s(z_{s,1})a}\right)\left(\frac{\chi_s(z_{s,2})-\chi}{\chi_s(z_{s,2})a}\right) P_m\left(k=\frac{l+0.5}{\chi}, z(\chi)\right) \end{aligned} \tag{3.7}$$

where $P_m(k,z)$ is the 3D matter power spectrum. We have assumed that $p_1(z)$ and $p_2(z)$ are normalized. $p_1(z)$ and $p_2(z)$ will be the same in the case of auto power spectrum, and different for the cross power spectrum. In figure 11 we show the theoretical convergence power spectrum calculated using halofit nonlinear matter power spectrum [23], as well as the power spectrum we measure using the simulated convergence map. After the PGD correction the power spectrum matches the halofit predictions.

4 Conclusions

In this paper we improve the halo statistics and small scale matter distribution in low resolution fast quasi N-body simulations. For halos, we introduce relaxed-FoF, a modification to the standard FoF algorithm so that the linking length is a function of the halo mass. For smaller halos, relaxed-FOF increases the linking length to enhance the identification of small halos, to improve agreement on the halo mass function, and to reduce the fraction of missed halos. We reject fake halos by reducing the linking length for the halos with large velocity dispersions. The rejection procedure removes fake halos found in the high density regions, and therefore improves the halo bias. We find that using a high resolution mesh for the 2LPT initial condition enhances the identification of small halos. We also calibrate the small-scale velocities of small halos by adding an irrotational velocity term. This extra term is written as the gradient of the velocity potential, which is learned from the gravitational potential. We verify the results on several halo statistics, including halo bias, halo mass function, halo auto power spectrum in real space and in redshift space, cross correlation coefficient with

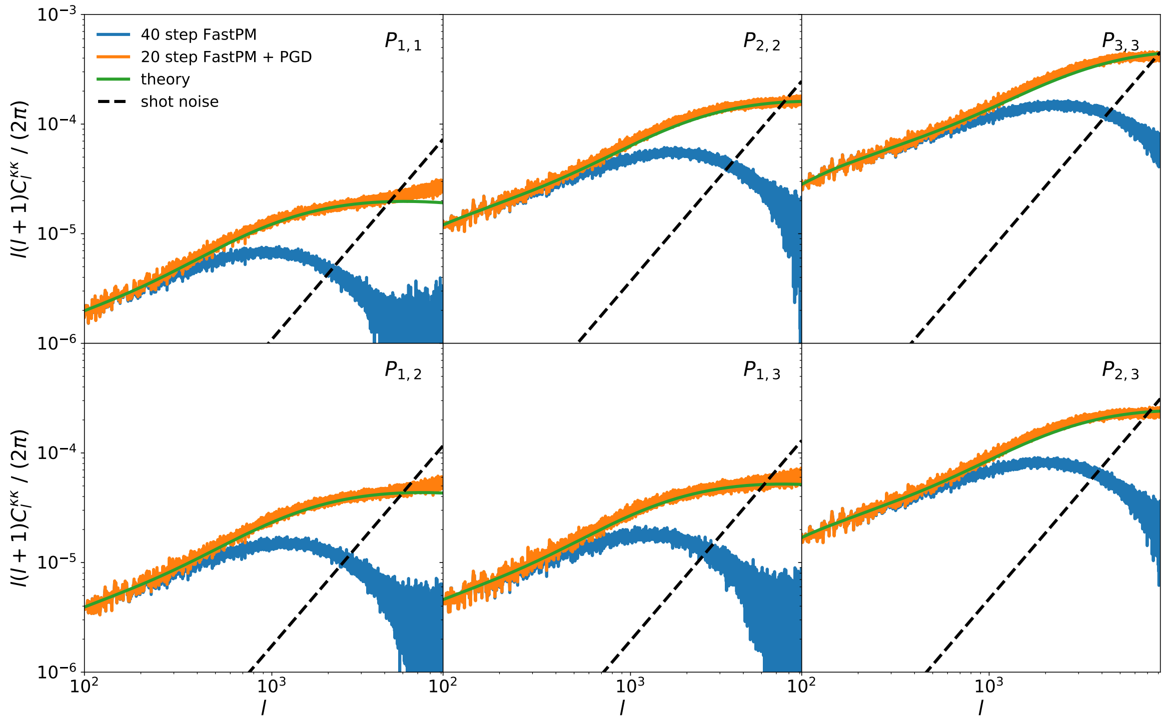

Figure 11. The weak lensing convergence auto power spectrum (upper panel) and cross power spectrum (bottom panel). The bin number 1, 2 and 3 corresponding to tomographic bin $z \in [0, 0.7]$, $[0.7, 1.4]$ and $[1.4, 2.1]$, respectively. The theoretical weak lensing convergence power spectrum is calculated using equation (3.7) with halofit nonlinear matter power spectrum. Black dashed lines show the shot noise power spectrum from the particles, which we subtract from the measured power spectrum.

the reference halo catalog, and halo-matter cross power spectrum. We find that our relaxed-FoF halo finder improves all of these. The ratio of missed halos and the halo catalog cross correlation coefficient suggest that our halo catalog from FastPM is comparable to the halo catalog from a full N-body simulation of the same mass resolution, while our catalog has better large scale auto power spectrum in real space and redshift space, as well as better halo-matter cross power spectrum.

We also incorporate the potential gradient descent (PGD) method into FastPM simulation to improve the matter distribution at nonlinear scales. We couple the PGD correction into the FastPM time steps. We show that the fully coupled PGD correction improves the matter power spectrum measured from static snapshot at all redshifts, just as the previously studied static PGD method [10]. We build a light-cone simulation from a PGD-enabled FastPM simulation, by interpolating the particles location between the steps. We show that the PGD correction significantly improves the convergence tomographic power spectrum measured from the light-cone output.

There are several free parameters in relaxed-FOF and PGD. In principle, these free parameters depend on simulation resolutions, the number of steps, and potentially cosmological parameters. To achieve the best results, they need to be optimized for different situations. One could obtain the parameters by fitting to a small volume high resolution simulation with the same random seed. Here we do not try to study the parameter dependence besides the redshift dependence, which contains the dominant effect of amplitude dependence. While we expect the dependence on the cosmological parameters other than amplitude to be small, this needs to be verified explicitly and is beyond the scope of this paper.

We plan to use FastPM for mock catalogs of both spectroscopic surveys such as DESI and photometric/weak lensing surveys such as LSST. The techniques we developed here will be useful to improve the halo and matter statistics in those simulations, thus enabling one to simulate the whole survey at the required mass resolution. For example, for DESI one needs to resolve halos down to $10^{11}h^{-1}M_{\odot}$ and to cover the entire survey one needs volumes in excess of $(3\,h^{-1}\,\mathrm{Gpc})^3$, which can be achieved with 10^{12} particle FastPM simulations, similar to the one that has recently been run [24].

Acknowledgments

The majority of the computation were performed on NERSC computing facilities Edison and Cori, billed under the cosmosim and m3058 repository. National Energy Research Scientific Computing Center (NERSC) is a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. We thank Dylan Nelson and the IllustrisTNG team for kindly providing the linear power spectrum and random seed of the IllustrisTNG simulations. This material is based upon work supported by the National Science Foundation under Grant Numbers 1814370 and NSF 1839217, and by NASA under Grant Number 80NSSC18K1274.

A Halo power spectrum of higher mass thresholds

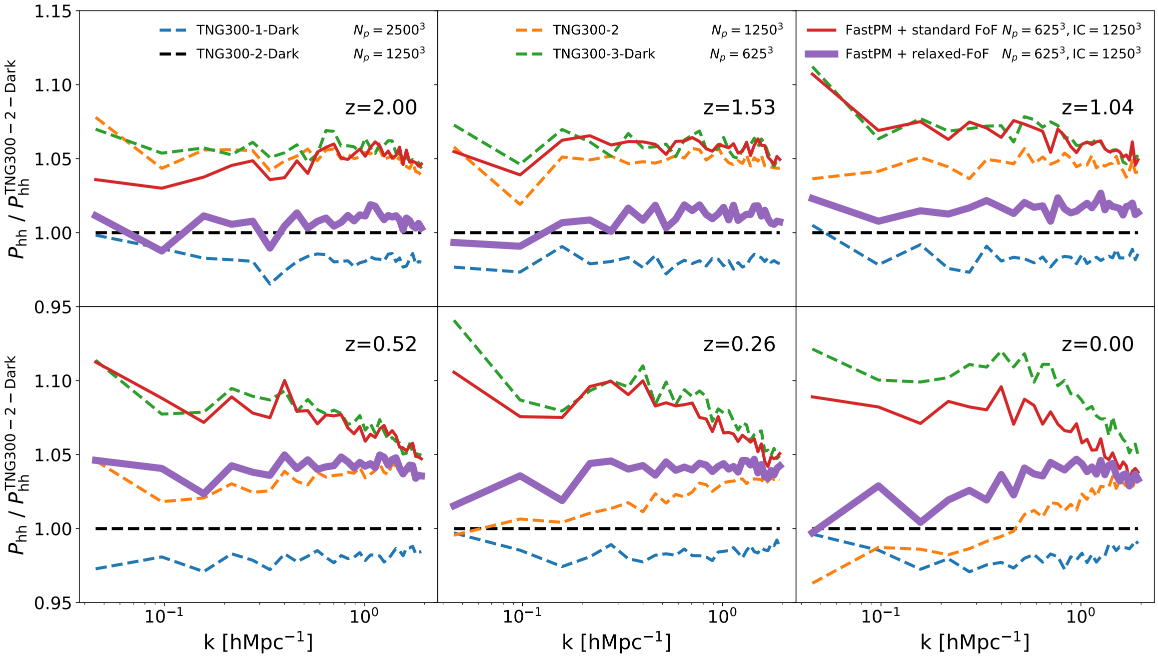

Figure 12. The ratio of halo auto power spectrum in real space from different simulations and various redshifts. The halos are selected using abundance matching, corresponding to $M \geq 2 \times 10^{11} M_\odot = 1.35 \times 10^{11} h^{-1} M_\odot$ halos (45 particles in FastPM).

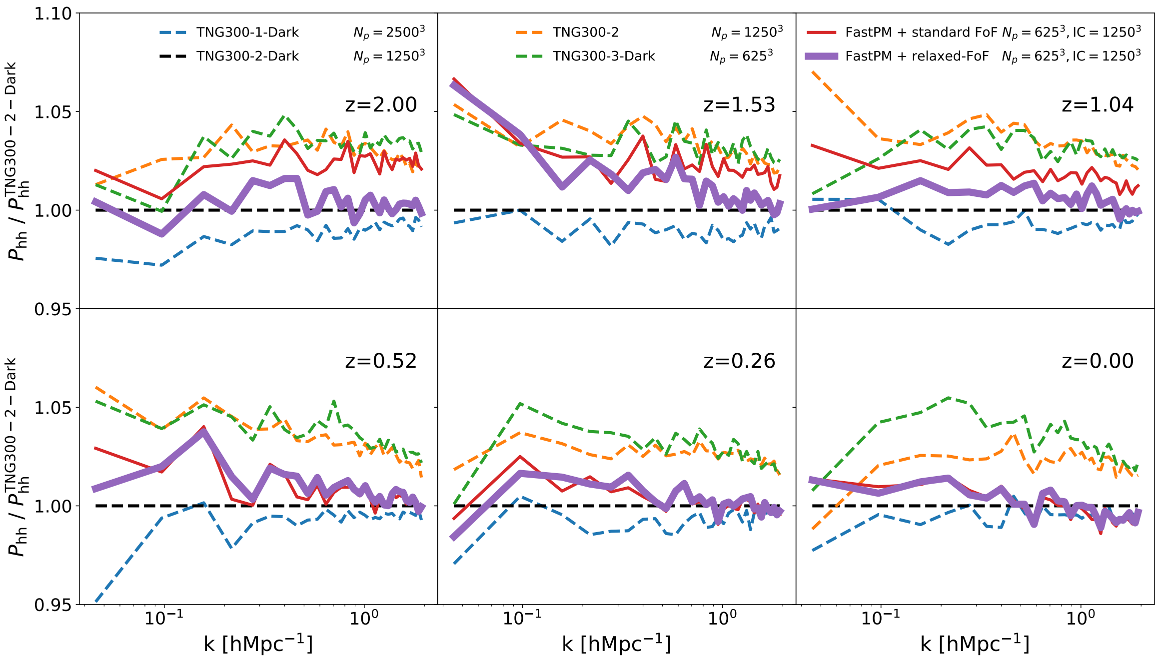

Figure 13. The same as figure 12, but for halos $M \geq 10^{12} M_\odot = 6.8 \times 10^{11} h^{-1} M_\odot$. For this mass range (200 particle halos) we expect the predictions from FastPM to be accurate, and the difference between relaxed-FoF and standard FoF should be small.

JCAP04(2020)002

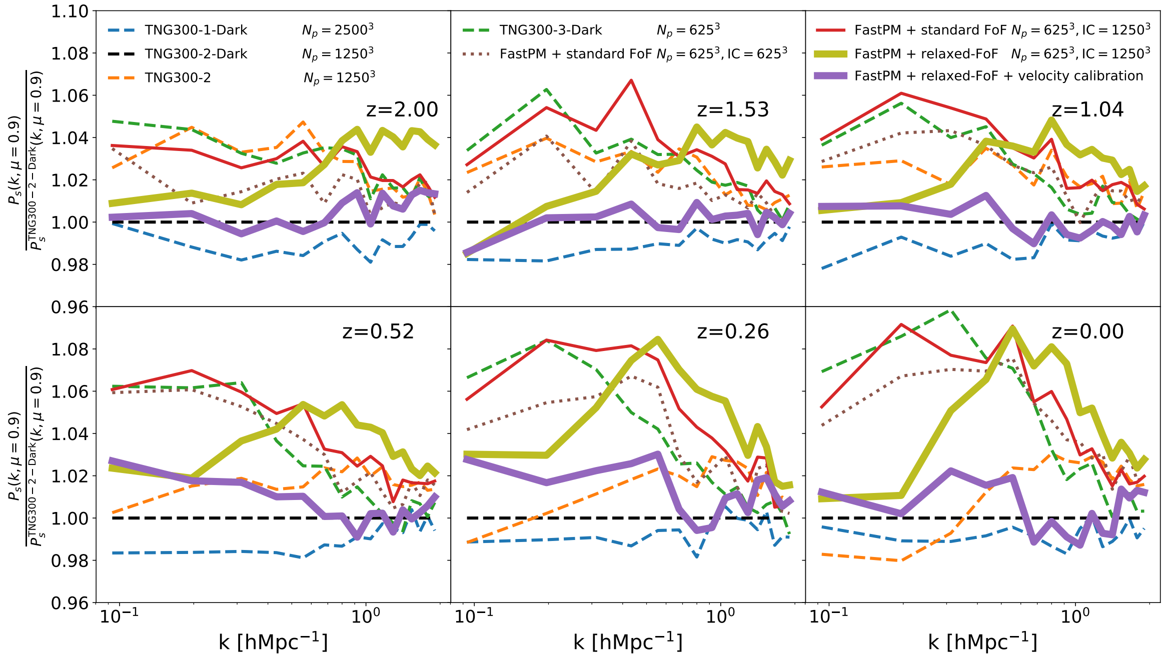

Figure 14. The ratio of halo auto power spectrum in redshift space from different simulations and various redshifts, for halos $M \geq 2 \times 10^{11} M_\odot = 1.35 \times 10^{11} h^{-1} M_\odot$. For RSD halo power spectrum with $\mu = 0$ see figure 12.

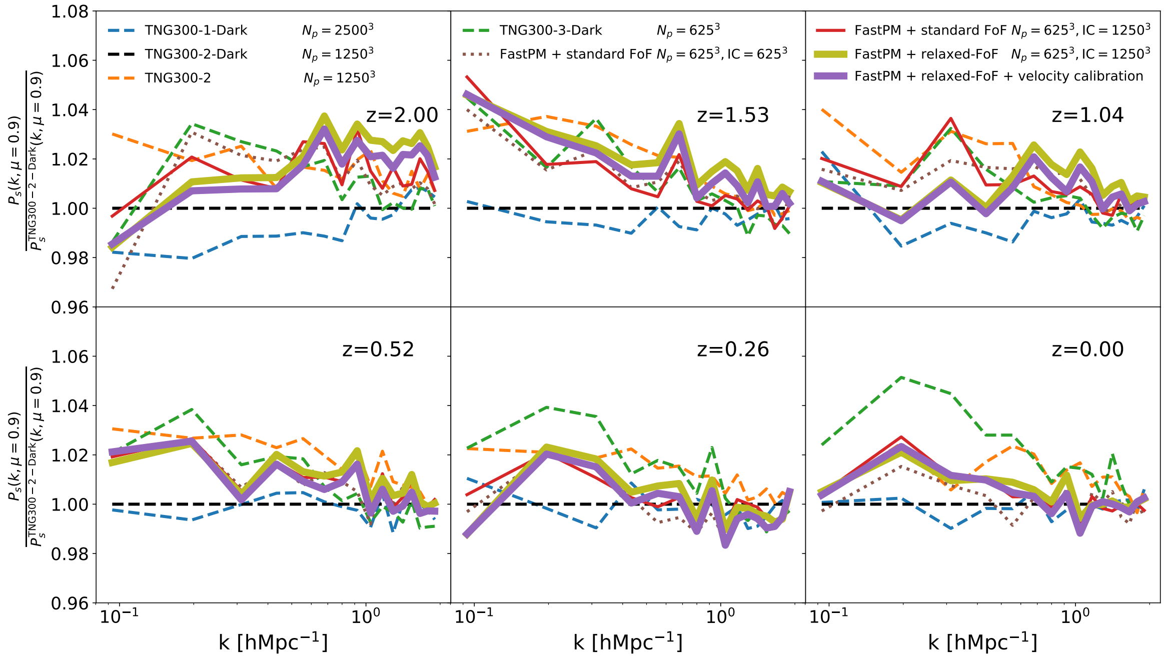

Figure 15. The same as figure 14, but for halos $M \geq 10^{12} M_\odot = 6.8 \times 10^{11} h^{-1} M_\odot$. For RSD halo power spectrum with $\mu = 0$ see figure 13. Again, we expect the predictions from FastPM to be accurate for this mass range (200 particle halos), and the difference between relaxed-FoF and standard FoF should be small.

JCAP04(2020)002

References

[1] F.-S. Kitaura et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: mock galaxy catalogues for the BOSS Final Data Release*, *Mon. Not. Roy. Astron. Soc.* **456** (2016) 4156 [arXiv:1509.06400] [INSPIRE].

[2] S. Joudaki et al., *KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **474** (2018) 4894 [arXiv:1707.06627] [INSPIRE].

[3] DES collaboration, *DES Y1 Results: Validating Cosmological Parameter Estimation Using Simulated Dark Energy Surveys*, *Mon. Not. Roy. Astron. Soc.* **480** (2018) 4614 [arXiv:1803.09795] [INSPIRE].

[4] Y. Feng, M.-Y. Chu, U. Seljak and P. McDonald, *FastPM: a new scheme for fast simulations of dark matter and haloes*, *Mon. Not. Roy. Astron. Soc.* **463** (2016) 2273 [arXiv:1603.00476] [INSPIRE].

[5] S. Tassev, M. Zaldarriaga and D. Eisenstein, *Solving Large Scale Structure in Ten Easy Steps with COLA*, *JCAP* **06** (2013) 036 [arXiv:1301.0322] [INSPIRE].

[6] DESI collaboration, *The DESI Experiment, a whitepaper for Snowmass 2013*, arXiv:1308.0847 [INSPIRE].

[7] LSST Science and LSST Project collaborations, *LSST Science Book, Version 2.0*, arXiv:0912.0201 [INSPIRE].

[8] DESI collaboration, *The DESI Experiment. Part I. Science,Targeting and Survey Design*, arXiv:1611.00036 [INSPIRE].

[9] J. DeRose et al., *The Aemulus Project I: Numerical Simulations for Precision Cosmology*, *Astrophys. J.* **875** (2019) 69 [arXiv:1804.05865] [INSPIRE].

[10] B. Dai, Y. Feng and U. Seljak, *A gradient based method for modeling baryons and matter in halos of fast simulations*, *JCAP* **11** (2018) 009 [arXiv:1804.00671] [INSPIRE].

[11] V. Springel et al., *First results from the IllustrisTNG simulations: matter and galaxy clustering*, *Mon. Not. Roy. Astron. Soc.* **475** (2018) 676 [arXiv:1707.03397] [INSPIRE].

[12] A.E. Evrard et al., *Virial Scaling of Massive Dark Matter Halos: Why Clusters Prefer a High Normalization Cosmology*, *Astrophys. J.* **672** (2008) 122 [astro-ph/0702241] [INSPIRE].

[13] M. Manera et al., *The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: a large sample of mock galaxy catalogues*, *Mon. Not. Roy. Astron. Soc.* **428** (2012) 1036 [arXiv:1203.6609] [INSPIRE].

[14] A.A. Klypin, S. Gottlöber, A.V. Kravtsov and A.M. Khokhlov, *Galaxies in N body simulations: Overcoming the overmerging problem*, *Astrophys. J.* **516** (1999) 530 [astro-ph/9708191] [INSPIRE].

[15] P.S. Behroozi, R.H. Wechsler and H.-Y. Wu, *The Rockstar Phase-Space Temporal Halo Finder and the Velocity Offsets of Cluster Cores*, *Astrophys. J.* **762** (2013) 109 [arXiv:1110.4372] [INSPIRE].

[16] J. Diemand, M. Kuhlen and P. Madau, *Early supersymmetric cold dark matter substructure*, *Astrophys. J.* **649** (2006) 1 [astro-ph/0603250] [INSPIRE].

[17] M. Maciejewski, S. Colombi, V. Springel, C. Alard and F.R. Bouchet, *Phase-space structures II: Hierarchical Structure Finder*, *Mon. Not. Roy. Astron. Soc.* **396** (2009) 1329 [arXiv:0812.0288] [INSPIRE].

[18] N. Hand, Y. Feng and C. Modi, *bccp/nbodykit: nbodykit v0.2.9*, (2017) https://doi.org/10.5281/zenodo.1051244.

[19] H. Hoekstra and B. Jain, *Weak Gravitational Lensing and its Cosmological Applications*, *Ann. Rev. Nucl. Part. Sci.* **58** (2008) 99 [arXiv:0805.0139] [INSPIRE].

[20] M. Kilbinger, *Cosmology with cosmic shear observations: a review*, *Rept. Prog. Phys.* **78** (2015) 086901 [arXiv:1411.0115] [INSPIRE].

[21] A. Zonca et al., *healpy: equal area pixelization and spherical harmonics transforms for data on the sphere in Python*, *J. Open Source Software* **4** (2019) 1298.

[22] K.M. Górski et al., *HEALPix — A Framework for high resolution discretization and fast analysis of data distributed on the sphere*, *Astrophys. J.* **622** (2005) 759 [astro-ph/0409513] [INSPIRE].

[23] R. Takahashi, M. Sato, T. Nishimichi, A. Taruya and M. Oguri, *Revising the Halofit Model for the Nonlinear Matter Power Spectrum*, *Astrophys. J.* **761** (2012) 152 [arXiv:1208.2701] [INSPIRE].

[24] C. Modi, E. Castorina, Y. Feng and M. White, *Intensity mapping with neutral hydrogen and the Hidden Valley simulations*, *JCAP* **09** (2019) 024 [arXiv:1904.11923] [INSPIRE].