



FASTPM: a new scheme for fast simulations of dark matter and haloes

Yu Feng,¹★ Man-Yat Chu,¹ Uroš Seljak^{1,2} and Patrick McDonald²

¹Berkeley Center for Cosmological Physics, Department of Physics, University of California Berkeley, Berkeley, CA 94720, USA

²Lawrence Berkeley National Laboratory, Berkeley, CA 94705, USA

Accepted 2016 August 22. Received 2016 August 19; in original form 2016 March 1

ABSTRACT

We introduce FASTPM, a highly scalable approximated particle mesh (PM) N -body solver, which implements the PM scheme enforcing correct linear displacement (1LPT) evolution via modified kick and drift factors. Employing a two-dimensional domain decomposing scheme, FASTPM scales extremely well with a very large number of CPUs. In contrast to Comoving-Lagrangian (COLA) approach, we do not require to split the force or track separately the 2LPT solution, reducing the code complexity and memory requirements. We compare FASTPM with different number of steps (N_s) and force resolution factor (B) against three benchmarks: halo mass function from friends-of-friends halo finder; halo and dark matter power spectrum; and cross-correlation coefficient (or stochasticity), relative to a high-resolution TREEPM simulation. We show that the modified time stepping scheme reduces the halo stochasticity when compared to COLA with the same number of steps and force resolution. While increasing N_s and B improves the transfer function and cross-correlation coefficient, for many applications FASTPM achieves sufficient accuracy at low N_s and B . For example, $N_s = 10$ and $B = 2$ simulation provides a substantial saving (a factor of 10) of computing time relative to $N_s = 40$, $B = 3$ simulation, yet the halo benchmarks are very similar at $z = 0$. We find that for abundance matched haloes the stochasticity remains low even for $N_s = 5$. FASTPM compares well against less expensive schemes, being only 7 (4) times more expensive than 2LPT initial condition generator for $N_s = 10$ ($N_s = 5$). Some of the applications where FASTPM can be useful are generating a large number of mocks, producing non-linear statistics where one varies a large number of nuisance or cosmological parameters, or serving as part of an initial conditions solver.

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

1 INTRODUCTION

Extracting full information from observations of the large-scale structure (LSS) of the universe, in weak lensing, galaxies, and other tracers, requires accurate predictions, which are only possible using simulations. These simulations can be used to create mock galaxy or weak lensing catalogues for covariance estimation (Knebe et al. 2015), to vary the predictions as a function of cosmological or nuisance parameters (e.g. galaxy formation parameters in halo occupation models as by Cooray & Sheth 2002), or even as a tool to generate initial conditions (Jasche & Wandelt 2013; Kitaura 2013; Wang et al. 2013). A full N -body simulation is too expensive both in terms of CPU-hours and wall-clock times. For this reason approximated N -body solvers that produce a reasonably accurate dark matter density field provides a practical alternative in critical

applications. They range from simple Lagrangian perturbation theory field realizations (Monaco et al. 2013; Kitaura et al. 2014) to N -body simulations optimized for specific applications.

A large class of codes that employ the latter is the particle mesh (PM) family (e.g. Merz, Pen & Trac 2005; Carlson, White & Padmanabhan 2009; Heitmann et al. 2010; White et al. 2010; Tashev, Zaldarriaga & Eisenstein 2013; White, Tinker & McBride 2014). The idea is to ignore the force calculation on small scales by skipping the TREE (PP) force part of a full TREEPM (P^3M) code. Vanilla PM has been criticized for failing to reproduce the linear theory growth at large scale as the number of time steps is reduced (see e.g. Izard, Crocce & Fosalba 2016). Recently, Tashev et al. (2013) introduced the Comoving-Lagrangian (COLA) enhanced PM scheme, where the large-scale displacement is governed by the analytic calculation from second-order Lagrangian perturbation theory (2LPT), and the PM is used only to solve for the ‘residual’ small-scale displacement that affects the formation of haloes. COLA scheme has gained a lot of attention recently (Tashev, Zaldarriaga & Eisenstein

*E-mail: yfeng1@berkeley.edu

2013; Howlett, Manera & Percival 2015; Koda et al. 2016; Tassev et al. 2015; Izard et al. 2016) because it adds a relatively small overhead to the vanilla PM method, yet enforces the correct growth on large scales. As expected, this approach fails on small scales, where approximate tree solvers (Sunayama et al. 2016) proposed to reduce the frequency of updating the small-scale force produce better results (percentage level agreement at high k), although at a higher computational cost. An alternative approach to reduce the cost of small-scale interactions is to neglect the TREE force calculation for particles that have formed haloes (Khandai & Bagla 2009).

The wall-clock time of a simulation is a somewhat overlooked issue (likely because the usual applications of approximate methods focus on a large number of independent mocks). The wall-clock time is still of relevance from a practical point of view, especially so for certain applications. For example, for Markov chains Monte Carlo or similar sampling algorithms, a shorter wall-clock time in the simulations allows more steps per chain which can be of significant importance. There are two possible ways to reduce the wall-clock time. The first is to redraw the compromise between amount of calculation and accuracy – usually, less accurate results can be obtained by faster simulations. The second approach is to employ more computing resources. In the second approach the idea is that the implementation of the scheme can efficiently use a larger amount of computing resource – the so-called strong scaling performance. Strong-scaling will become even more relevant as the number of computing nodes of supercomputing systems grows with time. Usually, the more complex is the code, the harder it is to enforce strong scaling, especially when moving to new platforms.

To address some of these issues, we implement a simple PM scheme into a new code we call FASTPM, where the linear theory growth of displacement (Zel'dovich approximation or 1LPT, where n LPT stands for Lagrangian perturbation theory at order n) is enforced by choosing an appropriate set of kick and drift factors. The PM solver in FASTPM is written from scratch to ensure that it scales extremely well with a large number of computing nodes. We implement it both within our approach, and within the COLA implementation, so that we can compare their performances.

We will describe the FASTPM code in Section 2 of this paper. We discuss the numerical schemes briefly in Section 2.1, then move on to discuss the two-dimensional parallel decomposition in Section 2.2 and show the strong scaling of FASTPM in Section 2.3. In Section 3, we explore the parameter space (number of time steps and force resolution) and investigate favourable schemes for approximated halo formation.

2 THE FASTPM CODE

2.1 Numerical schemes

The time integration in FASTPM follows the kick–drift–kick symplectic scheme described in Quinn et al. (1997), but is modified to achieve the correct large-scale growth. This is discussed further below. Here we discuss the transfer functions for force calculation in Fourier space.

In a PM solver, the gravitational force is calculated via Fourier transforms. First, the particles are painted to the density mesh, with a given a window function $W(r)$. We use a linear window of unity size (cloud-in-cell; Hockney & Eastwood 1988). We then apply a Fourier transform to obtain the overdensity field δ_k . A force transfer function relates δ_k to the force field in Fourier space.

There are various ways to write down the force transfer function $\nabla\nabla^{-2}$ in a discrete Fourier space. The topic has been explored

extensively by Hockney & Eastwood (1988) in the context of high-resolution PM and P³M simulations. FASTPM supports several combinations, which we describe below.

(i) Naive: naive Green's function kernels (k^{-2}) and differentiation kernels ($i\mathbf{k}$). This is used in COLA to solve for the residual motion with PM:

$$\nabla\nabla^{-2} = i\mathbf{k} k^{-2}. \quad (1)$$

(ii) H-E: sharpening naive scheme by de-convolving the cloud-in-cell (CIC) mass assignment window. This is an extremely popular set of choice because it is shown to minimize the small-scale error in force calculation,

$$\nabla\nabla^{-2} = i\mathbf{k} k^{-2} W_{\text{CIC}}^{-2}(k), \quad (2)$$

$$\text{where } W_{\text{CIC}} = \Pi_{d=x,y,z} \text{sinc}^2 \frac{k_d L}{2N}.$$

(iii) Finite differentiation kernel (FASTPM). The finite differentiation operator on a mesh is used, and no correction for mass assignment is applied. The three-point second-order central differentiation operator in Fourier space is (also mentioned in Hockney & Eastwood 1988)

$$\nabla^{-2} = \left(\sum_{d=x,y,z} \left(x_0 \omega_d \text{sinc} \frac{\omega_d}{2} \right)^2 \right)^{-1}, \quad (3)$$

where $x_0 = L/N_g$ is the mesh size, and $\omega = kx_0$ is the circular frequency that goes from $(-\pi, \pi]$, and

$$\nabla = D_1(\omega) = \frac{1}{6} (8 \sin \omega - \sin 2\omega). \quad (4)$$

D_1 is derived in Hamming (1989). The D_1 filter is exactly the same as the four-point discrete differentiation filter used in GADGET (Springel 2005), except we apply the function in Fourier space to simplify the implementation.

(iv) Gadget: naive Green's function kernels (k^{-2}), and with D_1 , applying sharpening corrections for mass assignment. This is the kernel used in GADGET for the PM part of the force:

$$\nabla\nabla^{-2} = iD_1(k) k^{-2} W_{\text{CIC}}^{-2}(k). \quad (5)$$

In solvers (e.g. TREEPM) where the short range force is calculated separately, a Gaussian function $W_g(k) = \exp[-(kr_s)^2]$ is often added to apodize the long-range force and suppress grid anisotropy, r_s being the smoothing scale of long-range force (typical slightly larger than the force mesh cell sizes) (Bagla 2002; Springel 2005; Habib et al. 2013). It is unclear this is necessary for pure PM simulations where there is no short-range force: some authors have completely neglected the smoothing filter, yet still obtained reasonable results (Tassev et al. 2013); authors in other fields have suggested that a high-order Gaussian smoothing performs well for certain problems (Hou & Li 2007). We find that even a mild smoothing, with $r_s = L/N_m$ of the mesh resolution, significantly reduces the number of haloes at the low-mass end; we will discuss the effect due to choice of schemes in Appendix A. For consistency in the main text paper we use the FASTPM finite differentiation scheme without smoothing, which as shown in Appendix A gives a sharper density field on larger scales, even though the kernels are less accurate on small scales in a traditional sense (Hockney & Eastwood 1988).

FASTPM accepts an arbitrary list of time steps. Shortcuts for several commonly used time stepping schemes (including linear and logarithmic) are provided.

(i) Time steps that are linear in scaling factor a . Linear stepping improves halo mass function, but requires assistance for an accurate

Table 1. Memory for the state vector per particle.

Column	Data type	Width (bytes)
x	Double	24
v	Single	12
a	Single	12
q	Integer	8
s_1	Single	12
s_2	Single	12
Total PM		56
Total COLA		80

linear scale growth factor if number of time steps is small (Tassev et al. 2013, 2015; Howlett et al. 2015; Izard et al. 2016).

(ii) Time steps that are linear in $\log a$. Logarithmic stepping improves growth on large (linear) scales at the cost of underestimating small scales and the halo mass function (White et al. 2014).

(iii) A hybrid scheme that controls the time step resolution in high redshift and low redshift independently:

$$(\delta a/a)^{-1} = \sqrt{(1/a_1)^2 + (a/a_2)^2}, \quad (6)$$

where a_1 controls the early time stepping and a_2 controls the late time stepping. An example value of $a_1 = 0.05$ and $a_2 = 0.025$ gives about 90 steps from $a = 0.025$ to 1.0 (Carlson et al. 2009).

In this paper, we will focus on linear a stepping.

To compensate for the lack of short-range resolution, the resolution of the force mesh can be boosted by a factor $B = N_m/N_g$. The size per side of the mesh (N_m) used for force calculation is B times the number of particles per side (N_g). Most recent authors of PM/COLA advocated a mesh of $B = 3$ in order to capture the non-linear formation of haloes (Tassev et al. 2013; Izard et al. 2016), while others advocated for $B = 1$ (Howlett et al. 2015). In FASTPM B can be provided as a function of time. We will investigate the choice of B in later in this paper.

2.2 Memory usage

FASTPM implements both COLA and vanilla PM. COLA requires storing the two 2LPT displacement fields s_1 and s_2 , inducing a memory overhead relative to vanilla PM. The state vector of a single particle in FASTPM contains the position, velocity, and acceleration, in addition, the initial position is encoded into an integer for uniquely identifying the particles. When COLA is employed, we also store the s_1 and s_2 terms. Currently, the position of particles x is stored in double precision to reduce systematic evolution of round-off errors near the edge of the boxes, which can be concerning if stored in single precision. One can also store the relative displacement from the original position in a single precision, which can further reduce the memory consumption by a few per cent.

The memory usage on the state vector is summarized in Table 1. The memory usage per particle is 56 bytes for vanilla PM and 80 bytes for COLA. To account for load imbalance and particles near domain surfaces (ghosts), we always overallocate the memory storage for particles by a factor of $A > 1$. Therefore, the total memory usage for storing the state vector is $M_1 = 56AN_g^3$ for PM, and $M'_1 = 80AN_g^3$ for COLA.

To avoid repeated conversion from particles to the mesh, FASTPM creates two copies of the force mesh. The memory usage for the force mesh is $M_2 = 2 \times 4(BN_g)^3 = 8B^3N_g^3$. Note that the buffer for domain decomposition and for creation of a snapshot is overlapped

with the force mesh, and it thus does not incur further memory allocation.

In summary, the total memory usage of FASTPM is

$$M = M_1 + M_2 = \begin{cases} (56A + 8B^3)N_g^3 & \text{PM,} \\ (80A + 8B^3)N_g^3 & \text{COLA.} \end{cases} \quad (7)$$

A is typically bound by $1 < A < 2$. Therefore, the additional memory cost of COLA relative to vanilla PM to store s_1 and s_2 is 37–40 per cent for $B = 1$, 20–27 per cent for $B = 2$, and 9–15 per cent for $B = 3$.

2.3 Domain decomposition

The domain decomposition in FASTPM is two-dimensional. Decomposing in two dimensions (resulting one-dimension ‘pencils’ or ‘stencils’) is an effective way to deploy large Fourier transforms on a massively parallel scale (see e.g. Pekurovsky 2012; Pippig 2013). Some gravity solvers implement a new two-dimensional Fourier transform library (e.g. Habib et al. 2013). We choose to reuse the publicly available implementation, PFFT by Pippig (2013) for its minimal design. We note that PFFT was also used to improve the scaling of p-GADGET by Feng et al. (2016).

FASTPM decomposes the particles into the same two-dimensional spatial domains of the real space Fourier transform mesh. In general, there are more mesh cells than number of particles (when $B > 1$), and it is therefore more efficient to create ghost particles on the boundary of a domain than creating ghost cells. We do not further divide the domain along the third dimension because that would induce further communication costs.

For large-scale computations, a two-dimensional decomposition has two advantages over one-dimensional decomposition: (1) a smaller total surface area; and (2) a more balanced load. The surface area is directly proportional to the amount of communication for ghost particles. The surface area increases linearly with the number of processes ($O[P]$) for one-dimensional decomposition; while for two-dimensional decomposition, the scaling is close to $O[P^{1/2}]$. The unit of parallelism in one-dimensional decomposition is a slab of $1 \times N_g \times N_g$. Therefore, when the number of processes is greater than the number of slabs ($P > N_g$), the Fourier transform becomes extremely imbalanced. The unit of parallelism in two-dimensional decomposition is a pencil of $1 \times 1 \times N_g$, and the constraint is $P > N_g^2$. For a typical mesh size of 8192 we used, the limit translates to $8192^2 = 67\,108\,864$ processors for two-dimensional decomposition, a limit cannot be reached even with the next generation exascale facilities.

We point out that FASTPM is not the first N -body solver implementing a two-dimensional domain decomposing scheme. Previous implementations (e.g. Habib et al. 2013; Feng et al. 2016) mostly focused on weak scaling of simulations to a large number of particles. The strong scaling of schemes that resolves galactic scale interaction (gravity, hydrodynamics, and feedback) typically suffer from the heavy imbalance as the average volume of a domain decreases, and requires overdecomposition of domains (Springel 2005; Menon et al. 2015). FASTPM does not usually suffer from the overdecomposition since it does not implement any small-scale interaction, as we will discuss in the next section.

2.4 Strong scaling of FASTPM

We run a series of tests to demonstrate the strong scaling performance of FASTPM on the Cray XC-40 system Cori at National

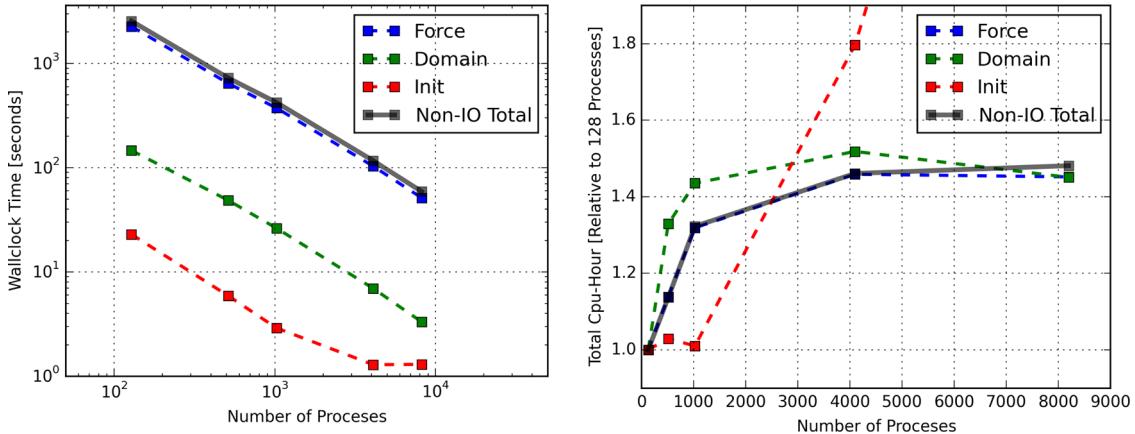


Figure 1. Strong scaling of FASTPM. We perform the test on a $1024 h^{-1} \text{Mpc}$ per side box on a 1024^3 particle grid. We run a total of 87 time steps (blend-space time stepping configuration) with the $B = 1, 2, 3$ force resolution scheme. We choose the variable force resolution scheme because in a traditional slab-based PM solver, the scaling would have stopped with 1024 cores for the $B = 1$ mesh. The three components shown are (1) force: the time spent in obtaining acceleration of particles; (2) domain: the time spent in migrating particles between different processors; (3) Init: the time spent in generating the 2LPT initial condition.

Energy Research Supercomputing Center (NERSC). The test simulation employs 1024^3 (1 billion) particles in a box of $1024 h^{-1} \text{Mpc}$ per side, and a resolution of $N_s = 40/B = 3$. The cosmology used is compatible to the *Wilkinson Microwave Anisotropy Probe (WMAP)* 9-year data.

The minimal number of computing nodes to run the simulation (due to memory constrains) is four, which translates to $N_p = 128$ computing cores. We then scale up the computing scale all the way up to $N_p = 8192$ (a factor of 64) computing cores. We measure the time spent in force calculation, domain decomposition, and the generation of the 2LPT initial condition. Time spent in file operations (IO) is not shown, since it follows closely to the status of the file system instead of the scaling of the code. We note, however, that the IO back-end of FASTPM (BIGFILE) is capable of achieving the peak performance of the file system in the BLUETIDES simulation (Feng et al. 2016).

We report the results of the scaling tests in Fig. 1. In the left-hand panel, we see that the scaling of the total wall-clock time is close to the ideal $1/N_p$ law. The scaling of 2LPT initial condition hits a plateau when more than 4096 computing cores are used, but this is not of particular concern since the fraction to the total only amounts 3 per cent even for the 8192 core run.

The deviation from the $1/N_p$ law is shown in the right-hand panel of Fig. 1. We show the evolution of the total CPU-hours (cost) as we increase the number of cores. For ideal $1/N_p$ scaling, the line should be flat. We see that the total cost increases slowly with the number of cores, by a factor of less than 1.45 when the number of cores increased by a factor of 64. The increase of total cost is primarily due to the imbalance of particles that is associated with the decrease of volume per domain. For example, with the 8192 core run, the most loaded domain handles three times of the average number of particles per domain. In a numerical scheme where small-scale force is fully resolved (TREPPM), the computing time increases very quickly with the growth of overdensity, and this imbalance would have significantly increased the cost. However, in an approximated PM solver (like FASTPM), the dominating cost component is Fourier transform, the load of which is balanced relatively well thank to the two-dimensional decomposition. The increase in synchronization time due to particle imbalance only slowly increase the total computing time. We point out that the

relatively quick increase in cost at small number of cores ($N_p < 1024$) is correlated with the crossing of communication boundaries of ‘local’/per cabinet network on the Cray XC 40 system Cori where the test is performed.

OPENMP threading is usually invoked as a workaround for strong-scaling limitations (e.g. Feng et al. 2016). For FASTPM this particular context is no longer relevant. In fact, running with multiple threads is always slower than running with equal number of processes, due to the lack of thread parallelism in the transpose phase of the parallel Fourier transform.¹ Therefore on current computer architectures, we do not recommend using more than one OPENMP thread in FASTPM.

2.5 Time stepping

When the number of time steps is limited, the kick–drift–kick integration scheme fails to produce the correct growth on large scales (e.g. Izard et al. 2016; Schneider et al. 2016). We note that even very large scales are not fully linear. As seen in Fig. 2, at $k_{\text{th}} \sim 0.03 h \text{Mpc}^{-1}$, the linear theory model introduces an average 0.5 per cent systematic error in mode amplitude at $z = 0$ due to non-linear coupling of modes (see also Takahashi et al. 2008; Baldauf, Mercolli & Zaldarriaga 2015; Foreman, Perrier & Senatore 2016; Seljak & Vlah 2015). The linear scale growth error is especially severe with linear a time stepping, where the relative change in the growth factor can be large – for example, if the first two time steps are at from $a = 0.1$ ($z = 9$) to $a = 0.2$ ($z = 4$). Our numerical experiment shows that the error is already 1.5 per cent with a single step from $a = 0.1$ to 0.2.

One way of eliminating this error without substantially increase the number of time steps is to insist the large-scale growth follows a model. For example, the COLA PM scheme calculates the large-scale trajectory of particles with 2LPT (Tassev et al. 2013), which yields an accurate growth of the large-scale modes. The disadvantages are higher memory requirement and the need to split the force into 2LPT and PM minus 2LPT, which requires tuning that is somewhat dependent on the number of time steps.

¹ Refer to <https://github.com/mpip/pfft/issues/6> for some discussions of this issue.

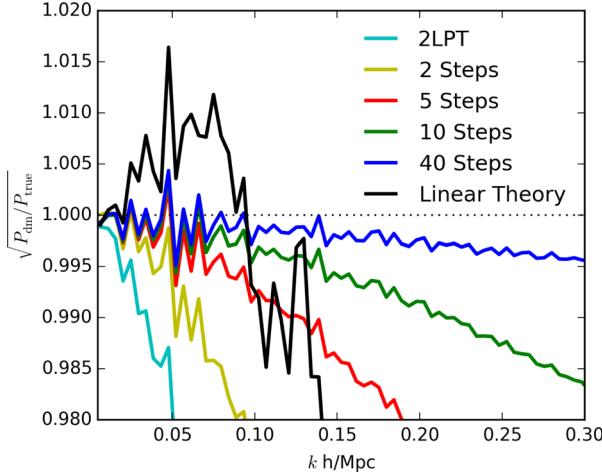


Figure 2. Recovery of linear growth with FASTPM. We show the recovery rate $\epsilon(k)$ at quasi-linear scales, divided by full N -body (RUNPB). Cyan: 2LPT; yellow: 2 steps; red: 5 steps; green: 10 steps; blue: 40 steps. The fluctuations are due to the sampling variance errors in recovering the initial condition (see text). We also show the ratio of linear modes to full N -body, which show that linear theory deviates from the non-linear results over most of the range.

In a pure PM scheme the error is due to the incorrect assumption that force and velocity remains constant during the course of a time step. This assumption is violated even in the linear regime. It can be compensated with a set of modified discrete drift and kick factors, motivated by the Zel'dovich (ZA) equation of motion. We derive next these factors (denoted with subscript FASTPM).

To see this, we first write down the Zel'dovich equation of motion to first order,

$$x_{\text{ZA}}(a) = q + D(a)s_1, \quad (8)$$

$$p_{\text{ZA}}(a) = a^2 \frac{dx_{\text{ZA}}}{dt} = a^2 \frac{dD}{da} \frac{da}{dt} s_1 \quad (9)$$

$$= \frac{dD}{da} a^3 E(a) s_1 \quad (10)$$

$$= a^3 E(a) g_p(a) s_1, \quad (11)$$

$$f_{\text{ZA}}(a) = a \frac{dp_{\text{ZA}}}{dt} = a^2 E(a) \frac{dp_{\text{ZA}}}{da} \quad (12)$$

$$= a^2 E(a) \frac{d[a^3 E g_p(a)]}{da} s_1 \quad (13)$$

$$= a^2 E(a) g_f(a) s_1. \quad (14)$$

We have followed the convention that $p = a^2 dx/dt$, $f = adp/dt$, and $E(a) = H(a)/H(a=1)$ is the dimensionless Hubble parameter. For simplicity we define the factors g_p and g_f as

$$g_p(a) = \frac{dD}{da}, \quad (15)$$

$$g_f(a) = \frac{d(a^3 E g_p)}{da} \quad (16)$$

$$= \frac{d(a^3 E)}{da} \frac{dD}{da} + \frac{d^2 D}{da^2} a^3 E(a) \quad (17)$$

$$= 3a^2 E \frac{dD}{da} + a^3 \frac{dE}{da} \frac{dD}{da} + \frac{d^2 D}{da^2} a^3 E, \quad (18)$$

and their integrals as

$$G_p(a) = D(a), \quad (19)$$

$$G_f(a) = a^3 E(a) g_p(a). \quad (20)$$

Next, we rearrange the ZA equation of motion as drift/kick operators by integrating ZA over the time step and eliminating the ZA displacement s_1 from the equation of motion,

$$\Delta[x_{\text{ZA}}]_{a_0}^{a_1} = \Delta[D(a)]_{a_0}^{a_1} s_1 \quad (21)$$

$$= \frac{p_{\text{ZA}}(a_r)}{a_r^3 E(a_r)} \frac{\Delta[G_p]_{a_0}^{a_1}}{g_p(a_r)}, \quad (22)$$

$$\Delta[p_{\text{ZA}}]_{a_0}^{a_1} = \frac{f_{\text{ZA}}(a_r)}{a_r^2 E(a_r)} \frac{\Delta[G_f]_{a_0}^{a_1}}{g_f(a_r)}. \quad (23)$$

We can now define the FASTPM modified drift and kick factor:

$$\mathcal{D}_{\text{FASTPM}} = \frac{\Delta[x_{\text{ZA}}]_{a_0}^{a_1}}{p_{\text{ZA}}} = \frac{1}{a_r^3 E(a_r)} \left(\frac{\Delta[G_p]_{a_0}^{a_1}}{g_p(a_r)} \right), \quad (24)$$

$$\mathcal{K}_{\text{FASTPM}} = \frac{\Delta[p_{\text{ZA}}]_{a_0}^{a_1}}{f_{\text{ZA}}} = \frac{1}{a_r^2 E(a_r)} \left(\frac{\Delta[G_f]_{a_0}^{a_1}}{g_f(a_r)} \right). \quad (25)$$

These operators can be used to construct any finite integration steps, which integrate the equation of motion of ZA solution exactly. Note that we keep a_r as the reference time for the step. In a standard kick–drift–kick scheme, a_r is a_0 at the first step, but otherwise a_r is between a_0 and a_1 .

It is trivial to show these factors converge to the usual drift and kick operators when the time steps are small. The usual drift and kick operators are (Quinn et al. 1997)

$$x(a_1) = x(a_0) + p(a_r) \int_{a_0}^{a_1} \frac{1}{a^3 E} da, \quad (26)$$

$$p(a_1) = p(a_0) + f(a_r) \int_{a_0}^{a_1} \frac{1}{a^2 E} da, \quad (27)$$

of which, the drift and kick factors are

$$\mathcal{D}_{\text{PM}} = \int_{a_0}^{a_1} \frac{1}{a^3 E} da, \quad (28)$$

$$\mathcal{K}_{\text{PM}} = \int_{a_0}^{a_1} \frac{1}{a^2 E} da. \quad (29)$$

When $a_1 \rightarrow a_0$, by definition we have

$$\frac{\Delta G_p}{\Delta a} \rightarrow g_p, \quad \frac{\Delta G_f}{\Delta a} \rightarrow g_f \quad (30)$$

and

$$\mathcal{D}_{\text{FASTPM}} \rightarrow \mathcal{D}_{\text{PM}}, \quad \mathcal{K}_{\text{FASTPM}} \rightarrow \mathcal{K}_{\text{PM}}. \quad (31)$$

Therefore, on infinitesimal time steps, the two sets of operators are identical.

In Fig. 2, we show the large-scale power spectrum of FASTPM divided by the full N -body simulation. We also show the comparison to the linear theory prediction, which deviates from non-linear already at very low k . Even with two time steps, i.e. a single full time step beyond 2LPT, the modified scheme is able to match the non-linear growth of a full N -body simulation at $k \sim 0.02 h \text{ Mpc}^{-1}$, a significant improvement over 2LPT or linear theory. In contrast, full N -body codes often do not match each other at very low k (Heitmann et al. 2008; Schneider et al. 2016), with constant offsets between them, likely due to inaccuracies in the linear growth. We

Table 2. List of simulations.

Model	Number of steps	Force resolution
COLA	5	3
PM	5	3
COLA	10	3
PM	10	3
COLA	10	2
PM	10	2
COLA	20	3
PM	20	3
COLA	40	3
PM	40	3

thus believe our modified kick–drift scheme could be useful even for standard high-resolution simulations.

We point out that it is possible (as we have done in the first version of this paper and `FASTPM`) to calibrate the kick or drift factor against linear theory, or non-linear power spectrum measured from a low-resolution PM simulation in the limit of many steps, which gives nearly identical results. Both time stepping schemes are also implemented in `FASTPM`.

3 DARK MATTER AND HALO BENCHMARKS WITH `FASTPM`

In this section, we investigate the accuracy of `FASTPM` against computational cost, varying several of its parameters. We focus particularly on halo formation with `FASTPM`, but we also compare it against the dark matter statistics. The particular application we have in mind is to find an approximate halo formation scheme that can be used for a fast extraction of galaxy statistics. To be more specific, we would like to find an approximation scheme that reduces the cost (CPU time), while also reducing the systematic error of the halo statistics to an acceptable level (defined more precisely below).

The parameters we investigate are number of time steps N_s and the force resolution $B = N_m/N_g$, the ratio between force mesh and number of particles per side. We also perform a comparison between (our version of) COLA and `FASTPM`. The `FASTPM` simulations used in this work are listed in Table 2.

The simulations are compared against a `TREPPM` simulation on 2048³ particles in a 1380 h^{-1} Mpc per side box (`RUNPB`). For the `FASTPM` simulations, we reconstruct the s_1 and s_2 2LPT terms from a single precision $z = 75$ initial condition of `RUNPB`, and extrapolate the initial displacement to $z = 9.0$. This ensures that the `FASTPM` simulations are using the same initial modes as `RUNPB`, up to the numerical errors. In all simulations, the friends-of-friends (FoF) finder uses a linking length of 0.2 times the mean separation of particles (Davis et al. 1985). Our results are robust against linking length, as the main purpose of linking length is to identify overdensity features and apply abundance matching. In Appendix C, we also compare the results of $N_s = 10/B = 2$ simulations against those from a shorter linking length of 0.168, which reduces the halo bridging effect of FoF.

We list the total CPU time used in each run in Fig. 3. The total CPU time increases with B and N_s . The most expensive scheme we considered is the $N_s = 40/B = 3$ scheme suggested by Izard et al. (2016), using 1300 CPU hours each on the reference system (70 times of 2LPT). Reducing the number of time steps and the force resolution can drastically reduce the computing time. For example, while $N_s = 40/B = 3$ costs 72 times 2LPT, our favourite scheme

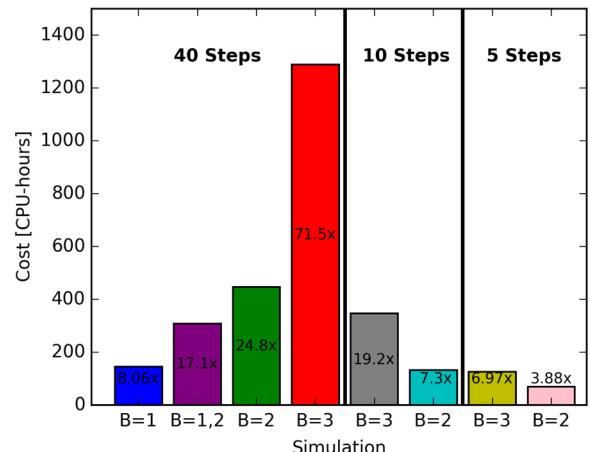


Figure 3. The integrated CPU-hours. The text in each vertical bar shows the ratio to generating a 2LPT initial condition. COLA and PM use almost same amount of CPU-hours, thus are not independently shown.

$N_s = 10/B = 2$ uses only 7 times of cost of 2LPT, while $N_s = 5/B = 2$ reduces this to 4.

3.1 Definitions of benchmarks

We use the ratio of mass function $\phi_1(M)/\phi_2(M)$ to illustrate the difference in mass function.² We use abundance matching to correct for the difference in mass function, but note that more complicated alternatives have been used by other authors as well (e.g. Sunayama et al. 2016; Izard et al. 2016).

We define the transfer function T as the square root of the ratio of the power spectra,

$$T(k, \mu) = \sqrt{P_1(k, \mu)/P_2(k, \mu)}. \quad (32)$$

The agreement is defined as good when T is close to 1. The transfer function measures the relative bias between two fields.

Note that transfer function is often defined as the ratio of the cross-power spectrum to auto-power spectrum: the two definitions are the same if the cross-correlation coefficient is unity. We define the cross-correlation coefficient r as

$$r(k, \mu) = P_{1,2}(k, \mu)/\sqrt{P_1(k, \mu)P_2(k, \mu)}, \quad (33)$$

where $P_{1,2}$ is the cross-power between the approximated and the accurate model. For haloes, we always use the catalogue after abundance matching them. We do this by rank ordering them by assigned halo mass, and then selecting the same number of the most massive haloes for the two catalogues. From this we can define another related quantity, the dimensionless stochasticity,

$$f(k, \mu) = \sqrt{(1 - n P_1(k, \mu))(1 - n P_2(k, \mu))} + (1 - n P_{1,2}(k, \mu)), \quad (34)$$

where n is the number density of haloes, which is identical in two simulations due to the abundance matching. The stochasticity is 1 when two catalogues contain completely different haloes, and 0 when two catalogues are identical. So stochasticity f expresses the fraction of misidentified haloes in the approximate simulation. Typically this happens because the code has assigned an incorrect mass to the halo, so that the halo does not enter the abundance matched catalogue. We could have also defined stochasticity using

² 1 stands for the approximated model and 2 stands for the accurate model.

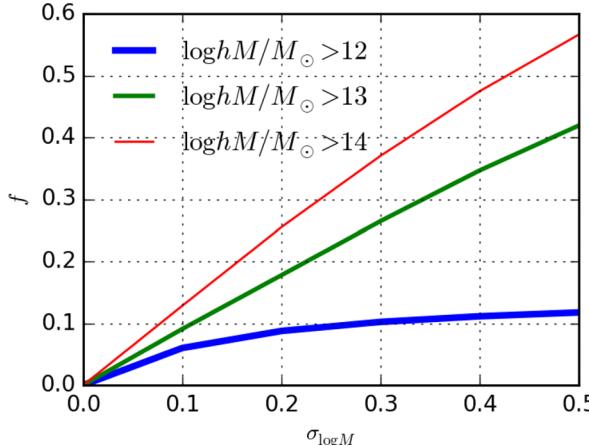


Figure 4. Stochasticity induced by mass scattering. We measure the stochasticity (large scale) due to adding mass scattering (given by σ) to the halo catalogue. Three different abundance cut corresponds to $M = 10^{12, 13, 14} h^{-1} M_\odot$ are shown in colours. The stochasticity increases with the mass scattering, which we identify as the main source of scale independent stochasticity (see text).

a mass error instead, but we do not pursue this here. Our definition of stochasticity is k dependent: if the halo positions are incorrect then we expect stochasticity f to increase with k , before decreasing again to converge to the shot noise limit.

While the numerical inaccuracies are one source of stochasticity, in practice we also have observational limitations. We typically observe galaxy luminosity, and select all the galaxies above a certain luminosity threshold (which can change with redshift). But there is a scatter between luminosity and halo mass and as a consequence a galaxy catalogue does not correspond to the most massive haloes at the same abundance (we are ignoring satellite galaxies in this discussion). To investigate this we take a given halo mass catalogue, add scatter to it, and re-rank the haloes. We then determine stochasticity, i.e. the fraction of haloes that are the same between this catalogue and the original one, at the same abundance.

In Fig. 4 we show the increase of stochasticity induced by increasing the mass scatter (in $\log M$) in the RUNPB halo catalogue. For example, the scatter of 0.23, as expected in Baryon Oscillation Spectroscopic Survey (BOSS) CMASS catalogue, corresponds to a stochasticity of 20 per cent for typical CMASS halo mass (Rodríguez-Torres et al. 2016). So as long as FASTPM stochasticity is significantly smaller than this there is no need to make it exactly zero.

It is worth pointing out that stochasticity or cross-correlation coefficient is in some sense the more important benchmark than the transfer function. This is because if the correlation coefficient is unity (or stochasticity zero) one can still recover the true simulation by multiplying the modes of the approximate simulation by the transfer function. Of course this requires the transfer function to be

known, but it is possible that it is a simple function of parameters, such that one can extract it from a small set of simulations without a major computational cost. In the following we will however explore all of the benchmarks defined here. These are summarized in Table 3 for each simulation.

3.2 Varying number of time steps

In this section, we discuss the effects due to varying the number of time steps employed in the simulation N_s , while fixing the force resolution at $B = 3$.

The transfer function and cross-correlation coefficient of matter density relative to the RUNPB simulation are shown in Fig. 5. With 5 steps, the cross-correlation coefficient is 93 per cent at $k = 1 h \text{Mpc}^{-1}$. Increasing the number of time steps does improve transfer functions and cross-correlation coefficients significantly. With 40 steps, the transfer function is close to 98 per cent at $k = 1 h \text{Mpc}^{-1}$, and the cross-correlation coefficient is close to 99 per cent at $k = 1 h \text{Mpc}^{-1}$. These metrics indicate that if a high-accuracy matter density field is of interest a 40 step simulation is preferred. It is worth pointing out that by extracting the transfer function, and then multiplying the modes with it, one obtains nearly perfect results up to $k = 1 h \text{Mpc}^{-1}$ even with 10 steps, given that the cross-correlation coefficient is 99 per cent or larger over this range.

We also observe that with the $N_s = 10$ runs COLA gives a larger transfer function than FASTPM at all scales, while $N_s = 5$ COLA gives a smaller transfer function and a significantly worse cross-correlation coefficients at all scales. It is worth noting that COLA has a free parameter n_{LPT} which needs to be tweaked for number of steps and cosmology parameters (Tassev et al. 2013). In our tests we find that COLA is very sensitive to the choice of this parameter. In contrast, the kick and drift scheme in FASTPM does not require tweaking of any parameters.

We next look at the haloes. Before applying abundance matching, we first show the mass function relative to RUNPB is shown in Fig. 6. For runs with more than 10 steps, the mass function has converged to 10 per cent agreement with RUNPB regardless of whether COLA or PM is used. However, if mass accuracy is required then 5 steps appears to be insufficient, as the 5 step PM run recovers only 80 per cent of the mass function (60 per cent for COLA). This is not necessarily a problem, since exact mass assignment is not required in a galaxy survey with a given abundance, as long as the rank ordering is preserved. However, in practice lower number of steps also introduces errors in the mass assignment, which increase the stochasticity, as we show below.

The rest of the benchmarks are calculated after applying abundance matching to reassign halo masses in the FASTPM simulations. In Fig. 7, we show the benchmarks on the transfer function, cross-correlation coefficient, and stochasticity of haloes as the total number of time steps N_s is varied. Three mass threshold

Table 3. List of benchmarks.

Samples	Transfer function ^a	Cross-correlation coefficient ^a	Stochasticity ^a	Mass function
Matter	Y	Y	N	N/A
$\log h^{-1}M/M_\odot \geq 12$	Y	Y	Y	Y
$\log h^{-1}M/M_\odot \geq 13$	Y	Y	Y	Y
$\log h^{-1}M/M_\odot \geq 14$	Y	Y	Y	Y

Note. ^aFor these benchmarks, the halo mass in FASTPM simulations are reassigned by abundance matching against haloes in RUNPB.

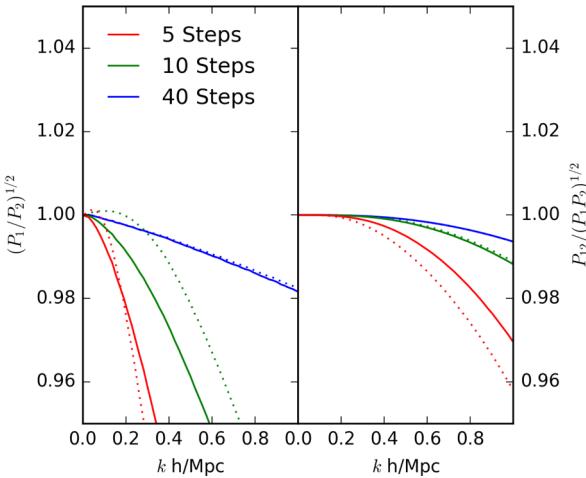


Figure 5. Benchmarks on matter density, varying number of time steps. Left: transfer function. Right: cross-correlation coefficient. Colours (online): $N_s = 5$ (red), 10 (green), and 40 (blue). Solid: PM; dots: COLA.

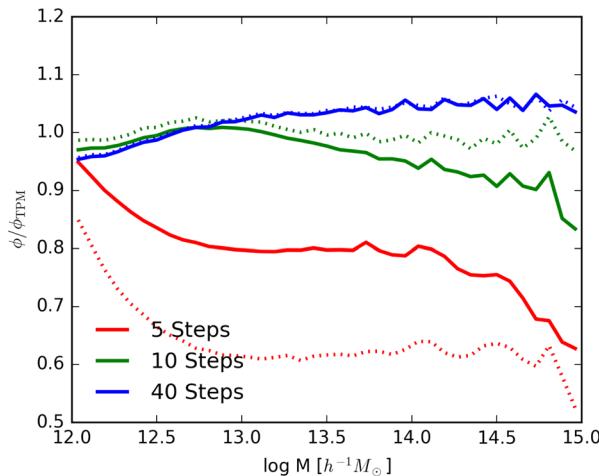


Figure 6. Benchmark of halo mass function, varying number of time steps. We show the number of time steps (colour: red, green, blue) from $N_s = 5$, 10, and 40. Solid: PM; dots: COLA.

$M = 10^{(12,13,14)} h^{-1} M_\odot$ are used. All the results are for $z = 0$. We find the following results.

(1) Beyond 10 steps the improvement is very limited. This is very different from the matter density field, where one gains major advantage using 40 steps. The additional steps therefore mostly improves the profile and velocity dispersion of haloes.

(2) For abundance matched haloes, PM out-performs COLA at low number of time steps. This can be most clearly seen in the 5 step runs, where PM is nearly a factor of 2 closer to exact solution at all k, μ values. The PM advantage over COLA decreases as the number of time steps increases. At 40 steps, PM and COLA converges to the same result. COLA splits the displacement into a residual field and a large-scale 2LPT component. It is worth noting that even though the 10 step COLA simulations give better matter transfer function than PM (as seen in Fig. 5), the advantage in matter density seems to be consistently hurting the performance in haloes. This could be because COLA has free parameters whose performance may have been optimized for the dark matter benchmarks. In contrast, there are no free parameters in FASTPM.

(3) FASTPM matches more massive haloes better than less massive haloes. For example, for the 10 step runs, the stochasticity reduces from 10 per cent at $M > 10^{12} h^{-1} M_\odot$ to 7 per cent at $M > 10^{14} h^{-1} M_\odot$. However, the overall stochasticity is given by f/n , and since n is a lot higher for the low-mass haloes the absolute stochasticity is a lot lower for low-mass haloes, despite the larger value of f . A 10 per cent stochasticity is likely more accurate than our current level understanding of the halo mass–galaxy luminosity relation (More et al. 2009; Yang, Mo & van den Bosch 2009; Behroozi, Conroy & Wechsler 2010). For example, the scatter in the halo mass at a fixed luminosity is 0.4 dex for $10^{12} h^{-1} M_\odot$ haloes (Behroozi et al. 2010), if all uncertainties are considered, and even larger for the larger halo masses. As a comparison, for a scatter of 0.18 dex in halo mass, we introduce a stochasticity of $f \sim 0.10, 0.18, 0.22$ for haloes of mass $10^{12,13,14} h^{-1} M_\odot$. Hence we believe that the stochasticity levels generated by $N_s = 10$, or even $N_s = 5$, suffice given the current observational uncertainties in halo mass determination.

(4) Redshift space statistics ($\mu > 0$, where $\mu = k_{||}/k$ and $k_{||}$ is the component of the wavevector k along the line of sight) are not very different from the real space statistics ($\mu = 0$). Hence redshift space distortions do not significantly affect the conclusions above.

(5) Scale dependence of f is weak for low-mass haloes, gradually increasing to higher masses. For the highest mass bin we observe it to increase by 0.05 to $k = 1 h \text{ Mpc}^{-1}$. We expect that a low-resolution PM gets the halo centres wrong by a fraction of their virial radius: hence, for lower mass haloes the absolute error is smaller.

Based on these observations one does not need more than 10 steps if haloes at $z = 0$ are of interest, and indeed for many applications even a 5 step FASTPM, possibly corrected with the transfer function, suffices. In addition, we comment that the requirement on the number of steps are similar at higher redshifts (we tested up to $z = 1$). Since our steps are uniform in expansion factor a , a 10 time step FASTPM simulation that runs to $z = 0$ would naively have the effective performance of 5 steps at $z = 1$, but we observe that actual performance is more in between 5 and 10 steps.

3.3 Varying force resolution B

In this section, we discuss the effects due to varying the resolution of the force mesh employed in the simulation B . As we have shown $N_s = 10$ is of sufficient accuracy for haloes, so we will fix the number of time steps at $N_s = 10$ in this section.

The transfer function and cross-correlation coefficient of dark matter density relative to the RunPB simulation are shown Fig. 8. We see that going from $B = 3$ to 1 the transfer function at $k < 1 h \text{ Mpc}^{-1}$ barely changes, and the cross-correlation coefficient changes even less.

As in the previous section, before applying abundance matching we investigate the mass function benchmark in Fig. 9. We see that regardless of whether PM or COLA is employed, as long as $B \geq 2$, the mass function is recovered at 90 per cent level for $M > 10^{12} h^{-1} M_\odot$. However, with a lower resolution, $B = 1$, only 80 per cent of the halo mass function is recovered at $M = 10^{13} h^{-1} M_\odot$, and even fewer haloes are found at lower masses. This indicates that due to the low resolution, the density contrast in $B = 1$ is insufficient for detecting haloes of $M \leq 10^{13} h^{-1} M_\odot$, as also seen in Lukić et al. (2007). However, one may still be able to salvage information about haloes in the regime where FoF finder fails. For example, it is possible to combine a stochastic sampling method (e.g. QPM or PATCHY;

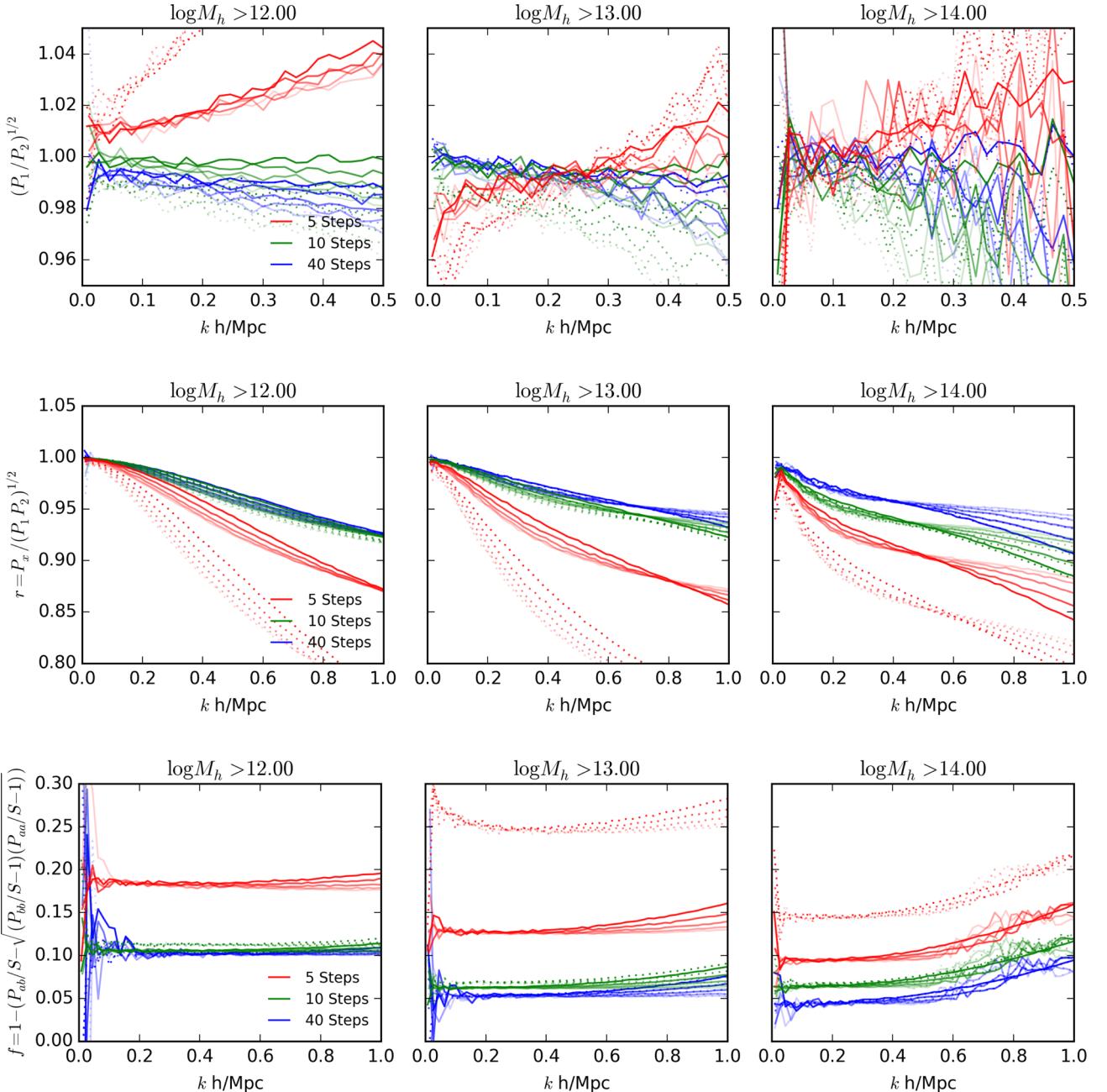


Figure 7. Benchmarks on haloes, varying number of time steps. Top panel: transfer function. Centre panel: cross-correlation coefficient. Bottom panel: stochasticity. Colours: $N_s = 5$ (red), 10 (green), and 40 (blue). Opacity: line of sight angle, $\mu = 0.1$ (transparent), 0.3, 0.5, 0.7, 0.9 (opaque). Solid: PM; dots: COLA. $S = 1/n$ is the shot-noise level.

Kitaura et al. 2014; White et al. 2014) for less massive haloes, but this may increase stochasticity and move f closer to unity.

The rest of the benchmarks are calculated after applying abundance matching to reassign halo masses in the FASTPM simulations. We show the rest of the benchmark suite in Fig. 10, which have been calculated after abundance matching. The structure of the figure is similar to Fig. 7, but now we vary the force resolution. We again observe some slight advantages of PM comparing to COLA (1 per cent level). The improvement due to increasing the force resolution from $B = 2$ to 3 is very limited, typically at 1 per cent level. The $B = 2$ approximation has a slightly larger (by 2 per cent) stochasticity than $B = 3$ approximation for the least massive threshold ($M > 10^{12} h^{-1} \text{M}_\odot$). Given that $B = 2$ simulation is almost three times

faster than a $B = 3$ simulation (Fig. 3), the 2 per cent increase in stochasticity is a reasonable price to pay. Overall, we find that the $N_s = 10/B = 2$ approximation uses about 10 per cent of CPU time of a $N_s = 40/B = 3$ simulation, yet the benchmarks on haloes of both are almost identical.

4 CONCLUSIONS

In this paper we introduce FASTPM, a new implementation of an approximate PM N -body solver. FASTPM modifies the standard kick and drift factors such that it agrees with Zel'dovich solution on large scales, guaranteeing zero error in $k \rightarrow 0$ limit even for very few time steps. These modified factors assume acceleration during the time

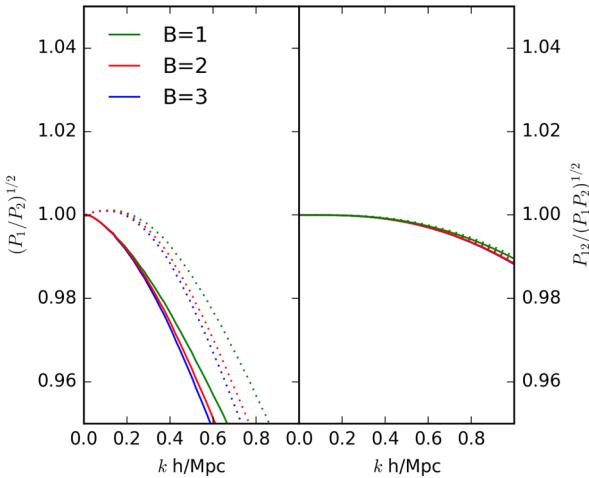


Figure 8. Benchmarks on matter density, varying force resolution. Red: $B = 2$; blue: $B = 3$; grey: $B = 1$. Solid: PM; dots: COLA. The number of steps is fixed to 10.

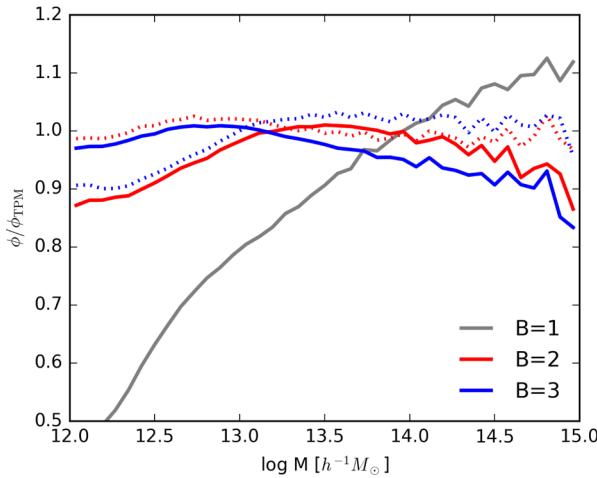


Figure 9. Benchmark on halo mass function, varying force resolution. Green: $B = 1$; red: $B = 2$; blue: $B = 3$. Solid: PM; dots: COLA. The number of steps is fixed to 10.

step is consistent with 1LPT, and they reduce to the standard ones in the limit of short time steps. We recommend the use of these modified factors whenever correct large-scale evolution is desired.

The domain decomposition in FASTPM for parallel Fourier Transform and particle data is in two dimensions, allowing the code to scale almost linearly with the number of CPUs when a large number of CPUs (more than 10 000) are employed.

We then proceed to investigate numerical precision of haloes identified with FoF within FASTPM. Four benchmarks are defined, measuring the performance: ratio of halo mass function, transfer function, cross-correlation coefficient, and stochasticity. We show that

(i) our implementation of PM (with modified kick and drift factors) performs slightly better than COLA on all benchmarks; especially when the number of time steps is low (5 steps);

(ii) the benchmarks on haloes of any scheme with $N_s \geq 10/B \geq 2$ is very close to the exact solution. This makes the $N_s = 10/B = 2$ approximation very interesting, as it uses only seven times of the computing time of generating a 2LPT initial condition, and 10 per cent of time of a $N_s = 40/B = 3$ approximated run. For higher

redshifts, or for cases where large stochasticity can be tolerated, the $N_s = 5/B = 2$ can also be adequate, at only four times the computing time of a 2LPT initial condition.

We see several use cases for FASTPM.

(i) As the halo catalogue step of a mock factory, FASTPM can be useful for generating a large number of mock catalogues. This is the same use case scenario similar to other recently proposed approximate N -body codes (Howlett et al. 2015; Sunayama et al. 2016; Izard et al. 2016).

(ii) Non-linear power spectrum (and higher order statistics) emulator: compared to other codes FASTPM can efficiently utilize a very large amount of computing resources. In fact, the turnaround time with FASTPM can be as low as 1 min if sufficient computing resources are reserved for real time usage. This means a large number of models, varying both cosmological parameters and nuisance parameters (such as halo occupation distribution parameters), can be explored rapidly. We plan to implement and expose a programming interface for various cosmology models in FASTPM.

(iii) Initial conditions solver: FASTPM is designed as a software library, making it easy to embed into another application. One option is the initial conditions solver (Wang et al. (2013)), which requires derivatives as a function of initial modes. For the derivatives to be computationally feasible one needs a simple force evaluation scheme and very few time steps, both satisfied by FASTPM. The scheme we have proposed, $B = 2$, $N_s = 10$, uses seven times more CPU time than a single 2LPT step, but provides realistic FoF haloes. Depending of the allowed stochasticity budget, using $N_s = 5$ or lower may also prove useful for computing the derivatives, given that the complexity of derivatives scales with N_s .

(iv) While we have focused on haloes in this paper, FASTPM can also be used for other applications, such as weak lensing (where dark matter is used). When it comes to dark matter our recommended strategy is to multiply the density field with the transfer function, which can be obtained from FASTPM itself ran at a higher resolution and number of time steps, or from a higher resolution simulation. Since the cross-correlation coefficient is very close to 1 up to $k = 1 \text{ h Mpc}^{-1}$, this would therefore guarantee high precision at least up to that k value. We expect the transfer function to be a slowly varying function of cosmological parameters and redshift. One possible strategy is to calibrate FASTPM against at sufficient number of points in parameter space to make this error negligible. Finally, we expect the loss of precision at even higher k to be related to the structure of haloes at small scales: FASTPM already finds all the haloes with the correct central position, and their mass error is relatively small, so only the internal halo profiles are incorrect. Since these are affected by baryon effects (gas profile, active galactic nuclei feedback, etc.) anyway this needs to be addressed in any pure dark matter code, and low-resolution FASTPM is not necessarily limited relative to a high-resolution N -body code. Another potential application is Lyman α forest, where a non-linear transformation of matter density can be used as an approximation to the Lyman α flux, with all three applications above being of possible interest. We plan to present some of these applications in the future.

ACKNOWLEDGEMENTS

We acknowledge support of NASA grant NNX15AL17G. The majority of the computing resources are provided at NERSC through the allocations for the Baryon Oscillation Spectroscopic Survey

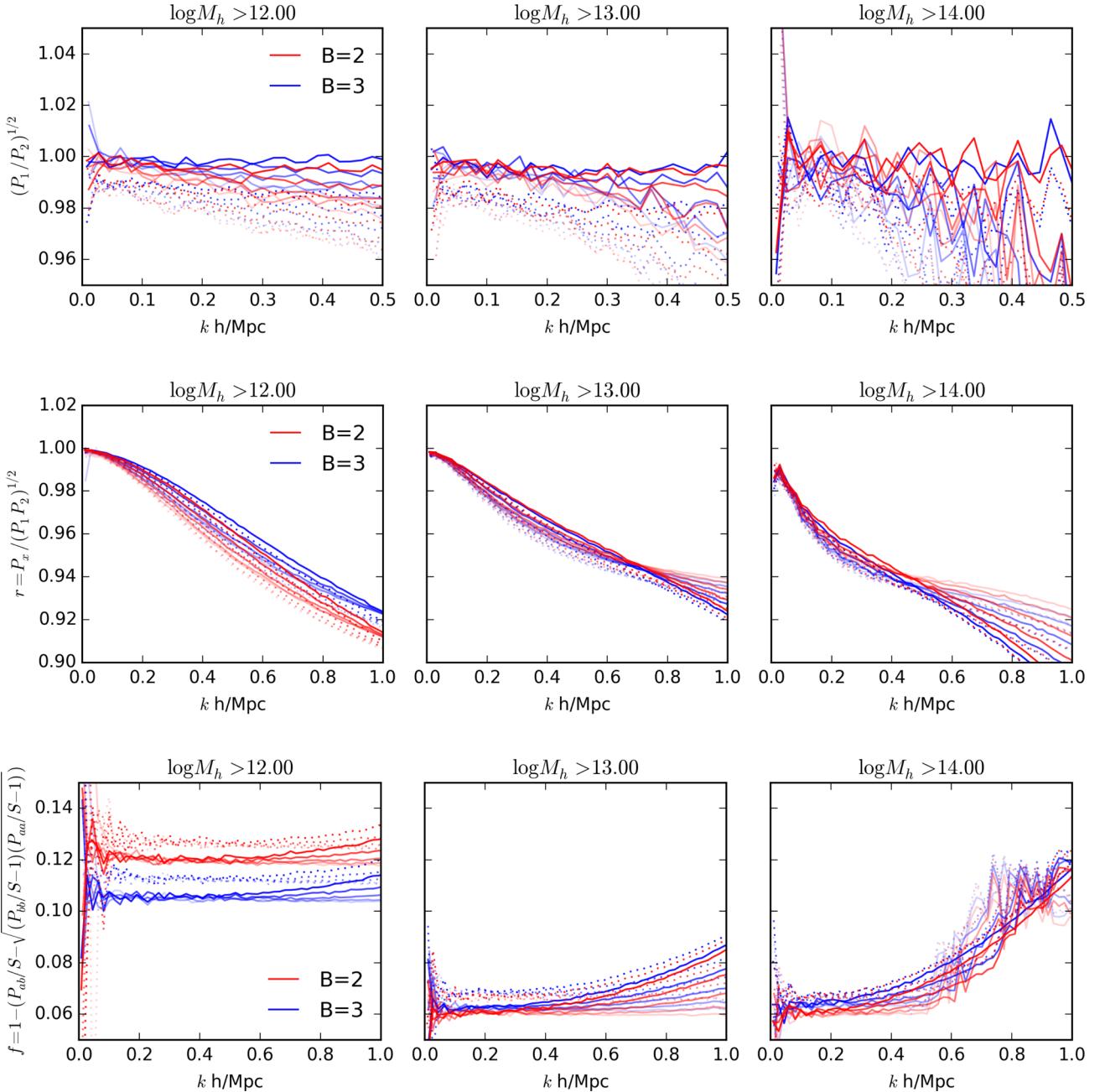


Figure 10. Benchmarks on haloes, varying force resolution. Top panel: transfer function. Centre panel: cross-correlation coefficient. Bottom panel: stochasticity. Red: $B = 2$; blue: $B = 3$; solid: PM; dots: COLA. Opacity: line of sight direction, $\mu = 0.1$ (transparent), 0.3, 0.5, 0.7, 0.9 (opaque). Solid: PM; dashed: COLA. $S = 1/n$ is the shot-noise level.

(BOSS) program and for the Berkeley Institute for Data Science (BIDS) program. We thank Dr Jun Kuda for distributing the source code of `COLA_HALO` under the GPLv3 license, which served both as a design inspiration and as a reference implementation of the COLA scheme. We thank Alejandro Cervantes and Marcel Schmittfull for their generous help in testing the code. We thank Martin White for providing the `RUNPB` `TREPPM` simulation and initial condition that formed the foundations of our benchmarks. The development of FASTPM is hosted by `github.com` at <https://github.com/rainwoodman/fastpm>. The version of code used in this paper is based on commit 9219d0. The data analysis software `NBODYKIT` is used for identifying FoF haloes and calculating of

power spectra.³ We welcome collaboration on development of both software packages.

REFERENCES

- Bagla J. S., 2002, *J. Astrophys. Astron.*, 23, 185
- Baldauf T., Mercolli L., Zaldarriaga M., 2015, *Phys. Rev. D*, 92, 123007
- Behroozi P. S., Conroy C., Wechsler R. H., 2010, *ApJ*, 717, 379
- Carlson J., White M., Padmanabhan N., 2009, *Phys. Rev. D*, 80, 043531

³ <https://github.com/bccp/nbodykit>, separate publication being prepared.

- Cooray A., Sheth R., 2002, Phys. Rep., 372, 1
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
- Feng Y., Di-Matteo T., Croft R. A., Bird S., Battaglia N., Wilkins S., 2016, MNRAS, 455, 2778
- Foreman S., Perrier H., Senatore L., 2016, J. Cosmol. Astropart. Phys., 5, 027
- Habib S., Morozov V., Frontiere N., Finkel H., Pope A., Heitmann K., 2013, in SC '13 Proceedings of SC13: International Conference on High Performance Computing, Networking, Storage and Analysis. ACM, New York, p. 6:1 (doi:10.1145/2503210.2504566, <http://doi.acm.org/10.1145/2503210.2504566>)
- Hamming R. W., 1989, Digital Filters, 3rd edn. Prentice-Hall, Hertfordshire, UK
- Heitmann K. et al., 2008, Comput. Sci. Discovery, 1, 015003
- Heitmann K., White M., Wagner C., Habib S., Higdon D., 2010, ApJ, 715, 104
- Hockney R. W., Eastwood J. W., 1988, Computer Simulation Using Particles. Taylor & Francis, Bristol, PA
- Hou T. Y., Li R., 2007, J. Comput. Phys., 226, 379
- Howlett C., Manera M., Percival W. J., 2015, Astron. Comput., 12, 109
- Izard A., Crocce M., Fosalba P., 2016, MNRAS, 459, 2327
- Jasche J., Wandelt B. D., 2013, MNRAS, 432, 894
- Khandai N., Bagla J. S., 2009, Res. Astron. Astrophys., 9, 861
- Kitaura F.-S., 2013, MNRAS, 429, L84
- Kitaura F.-S., Yepes G., Prada F., 2014, MNRAS, 439, L21
- Knebe A. et al., 2015, MNRAS, 451, 4029
- Koda J., Blake C., Beutler F., Kazin E., Marin F., 2016, MNRAS, 459, 2118
- Lukić Z., Heitmann K., Habib S., Bashinsky S., Ricker P. M., 2007, ApJ, 671, 1160
- Menon H., Wesolowski L., Zheng G., Jetley P., Kale L., Quinn T., Governato F., 2015, Comput. Astrophys. Cosmol., 2, 1
- Merz H., Pen U.-L., Trac H., 2005, New Astron., 10, 393
- Monaco P., Sefusatti E., Borgani S., Crocce M., Fosalba P., Sheth R. K., Theuns T., 2013, MNRAS, 433, 2389
- More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, MNRAS, 392, 801
- Pekurovsky D., 2012, SIAM J. Sci. Comput., 34, C192
- Pippig M., 2013, SIAM J. Sci. Comput., 35, C213
- Quinn T., Katz N., Stadel J., Lake G., 1997, ApJ, preprint ([astro-ph/9710043](http://arxiv.org/abs/astro-ph/9710043))
- Rodríguez-Torres S. A. et al., 2016, MNRAS, 460, 1173
- Schneider A. et al., 2016, J. Cosmol. Astropart. Phys., 4, 047
- Seljak U., Vlah Z., 2015, Phys. Rev. D, 91, 123516
- Springel V., 2005, MNRAS, 364, 1105
- Sunayama T., Padmanabhan N., Heitmann K., Habib S., Rangel E., 2016, J. Cosmol. Astropart. Phys., 5, 051
- Takahashi R. et al., 2008, MNRAS, 389, 1675
- Tassev S., Zaldarriaga M., Eisenstein D. J., 2013, J. Cosmol. Astropart. Phys., 6, 36
- Tassev S., Eisenstein D. J., Wandelt B. D., Zaldarriaga M., 2015, preprint ([arXiv:1502.07751](http://arxiv.org/abs/1502.07751))
- Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
- Wang H., Mo H. J., Yang X., van den Bosch F. C., 2013, ApJ, 772, 63
- White M., Pope A., Carlson J., Heitmann K., Habib S., Fasel P., Daniel D., Lukic Z., 2010, ApJ, 713, 383
- White M., Tinker J. L., McBride C. K., 2014, MNRAS, 437, 2594
- Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 695, 900

APPENDIX A: CHOICE OF DIFFERENTIATION SCHEME

FASTPM supports a variety of differentiation schemes. (Section 2). We explore the effect on dark matter density field due to choice of scheme in this section. This effect is shown in Fig. A1, where we compare the $z = 0$ power spectrum resulted from different schemes on the same initial condition with 10 and 40 step

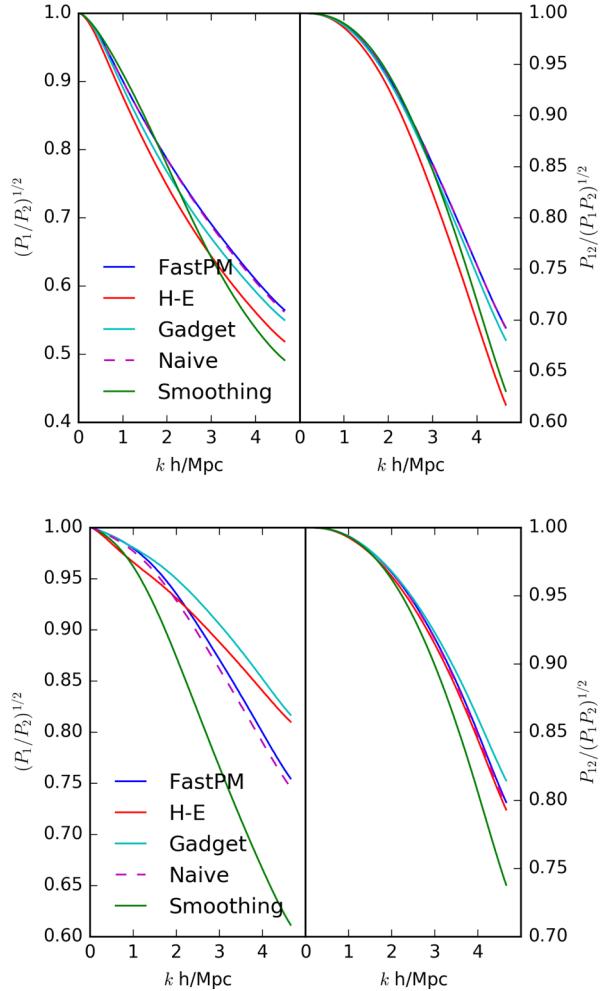


Figure A1. Comparing $z = 0$ power spectrum resulted from different schemes on the same initial condition with 10 and 40 step simulations. Upper panels: 10 steps; lower panels: 40 steps. The reference is the RUNPB simulation. The configuration of the simulations is the same as in Section 3, starting from $z = 9$.

simulations against the RUNPB simulation. The configuration of the simulations is the same as in Section 3, with a force mesh resolution of $L/N_m = 0.34 h^{-1}$ Mpc. Various schemes show overall a good level of agreement at $k < 0.5 h \text{ Mpc}^{-1}$. The difference increases to 5 per cent at $k = 2 h \text{ Mpc}^{-1}$. At 10 steps, the FASTPM kernel and the Naive kernel produce the largest power and the highest cross correlation coefficient. At 40 steps, the GADGET kernel shows improvements over FASTPM and Naive kernel at $k > 1 h \text{ Mpc}^{-1}$. Smoothing (Gaussian with $r_s = L/N_m$) slightly increases the cross-correlation coefficient at the cost of severely suppressing the power on small scales that corresponds to the collapse of haloes, losing haloes below $M < 10^{13.5} h^{-1} \text{ M}_\odot$ (see Fig. A2).

APPENDIX B: CHOICE OF STARTING REDSHIFT

Starting redshift affects the accuracy of simulations by two competing effects. We show this in Fig. B1, in Section B1.

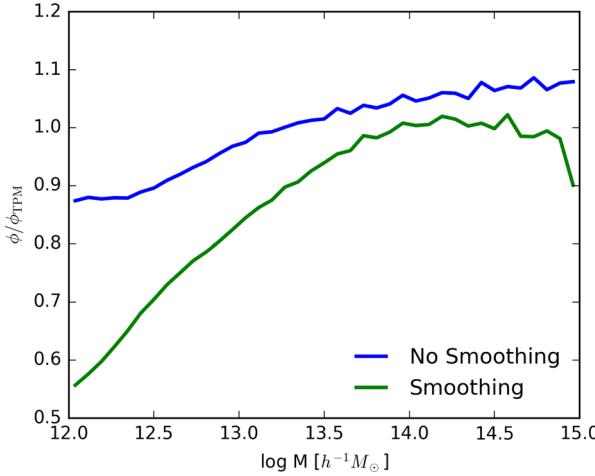


Figure A2. Comparing $z = 0$ mass function with and without smoothing on the same initial condition with 10 step simulations. The reference is the RUNPB simulation. Blue: FASTPM scheme without smoothing; green: with smoothing. The configuration of the simulations is the same as in Section 3.

(1) The temporal resolution decreases as the starting redshift increases. This effect is the most evident in the 10 step simulations, where the time steps are coarse. We see that the $z_i = 75$ simulation performed worse than $z_i = 9$ and 19. Therefore, for simulations with very few time steps, a lower starting redshift improves the benchmarks.

(2) The stochasticity in initial condition decreases as the starting redshift increases. In our case we want to minimize it against that of the reference simulation. We see this with the 200 steps simulation, where the $z_i = 75$ simulation, which is the starting redshift of reference TREEPM simulation, has the least error in power spectrum. The improvement due to better agreement in initial condition only affects the power spectrum, not the cross-correlation coefficient. Therefore, for simulations with many time steps, matching the starting redshift to that of the reference N -body simulation gives the best benchmarks. This is not surprising because the large-scale gravity force in the reference N -body simulation is calculated with the same PM method, although with many more time steps.

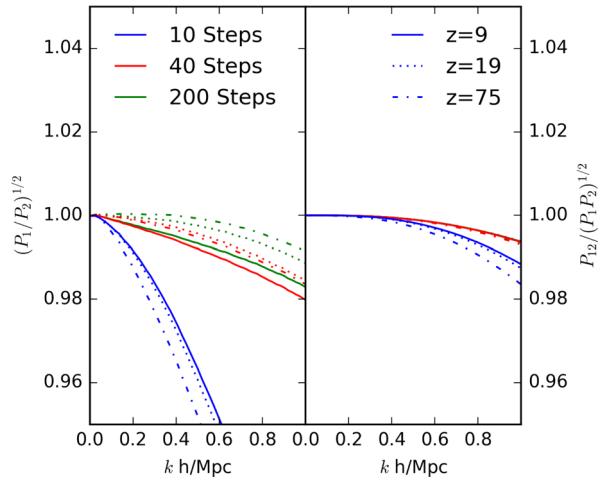


Figure B1. Comparing $z = 0$ power spectrum resulted from different starting redshifts on the same initial condition with 10 (blue), 40 (red), and 200 (green) step simulations. We show three starting redshifts, $z_i = 9$ (solid), 19 (dotted), and 75 (dotted-dashed). The reference is the RUNPB simulation, which started at redshift $z_i = 75$. Left: the square root of the ratio of power spectrum. Right: the cross-correlation coefficient. Note that in the right-hand panel the 200 step lines are covered by the 40 step lines.

The effect between $z_i = 19$ and 75 is about 0.4 per cent in the power spectrum at $k = 1 \text{ h Mpc}^{-1}$ at $z = 0$.

APPENDIX C: CHOICE OF LINKING LENGTH

The FoF algorithm that we use to identify objects are known to bridge haloes into the same object. The effect can affect our halo detection because in FASTPM the density field has less contrast than an accurate N -body simulation. Reducing linking length to 0.168 can alleviate this problem and improve the quality of the mock halo catalogue (Tinker et al. 2008). In Fig. C1 we show that reducing linking length from 0.2 to 0.168 does not affect the benchmarks, except for the least massive bin ($10^{12} h^{-1} \text{ M}_\odot$), where we observe a constant 2 per cent increase in bias and stochasticity.

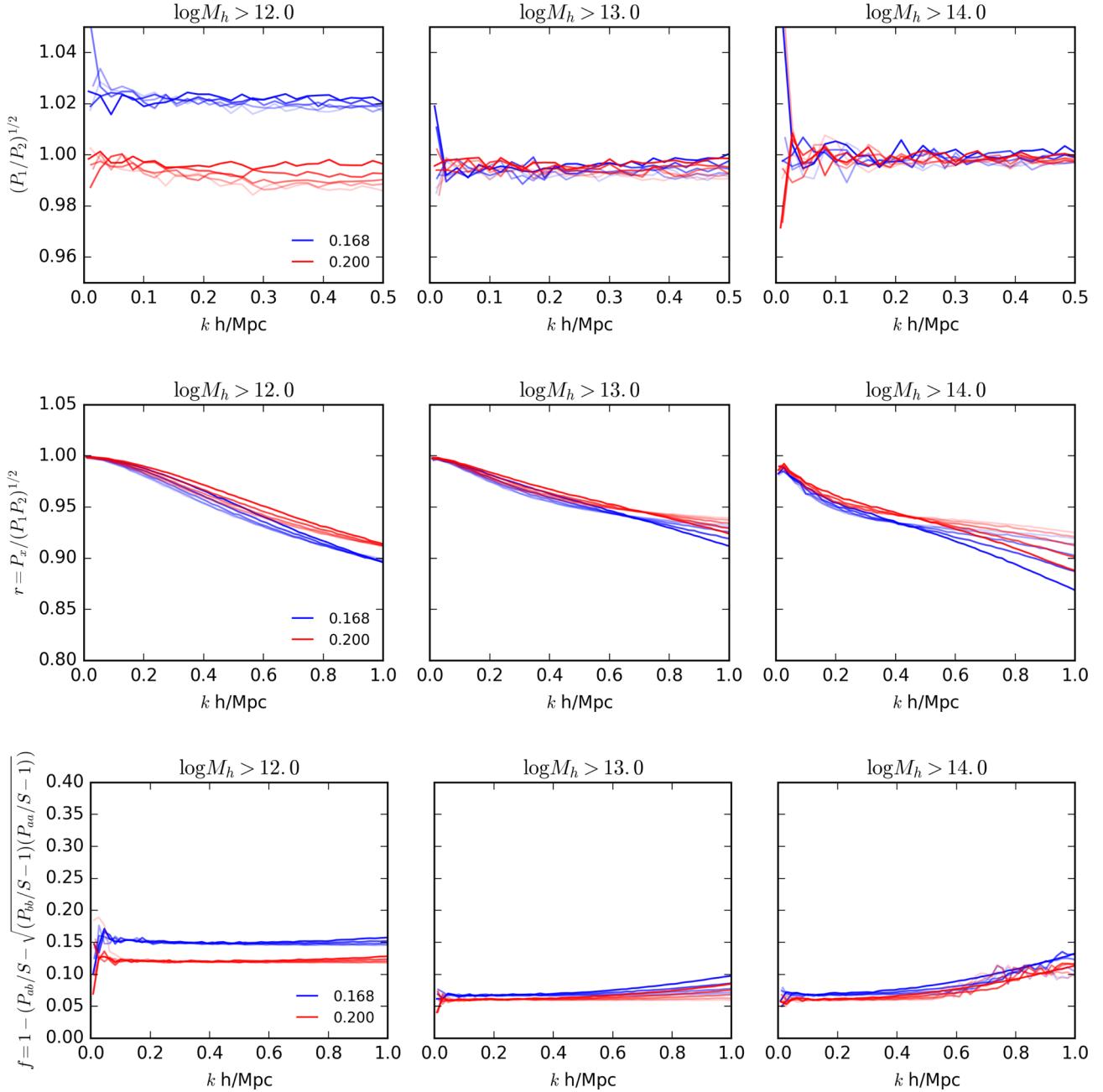


Figure C1. Benchmarks on haloes, varying linking length. Top panel: transfer function. Centre panel: cross-correlation coefficient. Bottom panel: stochasticity. Red: 0.200; blue: 0.168; opacity: line of sight direction, $\mu = 0.1$ (transparent), 0.3, 0.5, 0.7, 0.9 (opaque). $S = 1/n$ is the shot-noise level.

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