Particle initialization effects on Lyman-α forest statistics in cosmological SPH simulations

Nabendu Kumar Khan,[1][*] Girish Kulkarni,[1] James S. Bolton,[2] Martin G. Haehnelt,[3,4]
Vid Iršič,[4] Ewald Puchwein,[5] and Shikhar Asthana[3,4]

[1] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[2] *School of Physics and Astronomy, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[3] *Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[4] *Kavli Institute of Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*
[5] *Leibniz-Institut für Astrophysik Potsdam, An der Sternwarte 16, D14482 Potsdam, Germany*

ABSTRACT

Confronting measurements of the Lyman-α forest with cosmological hydrodynamical simulations has produced stringent constraints on models of particle dark matter and the thermal and ionization state of the intergalactic medium. We investigate the robustness of such models of the Lyman-α forest, focussing on the effect of particle initial conditions on the Lyman-α forest statistics in cosmological SPH simulations. We study multiple particle initialization algorithms in simulations that are designed to be identical in other respects. In agreement with the literature, we find that the correct linear theory evolution is obtained when a glass-like configuration is used for initial unperturbed gas particle positions alongside a regular grid configuration for dark matter particles and the use of non-identical initial density perturbations for gas and dark matter. However, we report that this introduces a large scale-dependent distortion in the one-dimensional Lyman-α transmission power spectrum at small scales ($k > 0.05$ s/km). The effect is close to 50% at $k \sim 0.1$ s/km, and persists at higher resolution. This can severely bias inferences in parameters such as the dark matter particle mass. By considering multiple initial conditions codes and their variations, we also study the impact of a variety of other assumptions and algorithmic choices, such as adaptive softening, background radiation density, particle staggering, and perturbation theory accuracy, on the matter power spectrum, the Lyman-α flux power spectrum, and the Lyman-α flux PDF. This work reveals possible pathways towards more accurate theoretical models of the Lyman-α forest to match the quality of upcoming measurements.

Key words: cosmology: theory – large-scale structure of Universe – intergalactic medium – methods: numerical

1 INTRODUCTION

The Lyman-α forest absorption features seen in high-redshift quasar spectra are the pre-eminent tool for probing the small-scale structure of the Universe. Properties of the absorption lines that constitute the Lyman-α forest are sensitive to the small-scale matter power spectrum as well as the thermal evolution of baryons. Furthermore, the forest is relatively insensitive to the non-linear baryonic physics at larger overdensities. Consequently, hydrodynamical cosmological simulations can be used in combination with Lyman-α forest measurements to put excellent constraints on various cosmological properties (see Meiksin 2009 and McQuinn 2016 for reviews).

Examples of constraints on the cosmological small-scale structure include constraints derived by various authors from the Lyman-α forest on dark matter models that affect the matter power spectrum at small scales. Multiple authors have inferred lower bounds on the warm dark matter particle mass

using Lyman-α forest statistics (Viel et al. 2005, 2008, 2013; Iršič et al. 2017a; Baur et al. 2017; Palanque-Delabrouille et al. 2020; Villasenor et al. 2023), with recent work reporting $m_x > 5.5$ keV (Iršič et al. 2023). Similar work has yielded constraints on the mass of a sterile neutrino (Viel et al. 2005, 2008; Baur et al. 2017; Garzilli et al. 2019) and fuzzy or ultra-light axion dark matter mass (Iršič et al. 2017b; Armengaud et al. 2017; Iršič et al. 2020; Rogers & Peiris 2021). Within the ΛCDM model, numerous authors have used the Lyman-α forest to constrain the thermal state of the IGM (Schaye et al. 2000; Ricotti et al. 2000; Zaldarriaga et al. 2001; Theuns et al. 2002; Lidz et al. 2010; Becker et al. 2011; Lee et al. 2015; Garzilli et al. 2017; Rorai et al. 2017a,b, 2018; Boera et al. 2019; Walther et al. 2019; Hiss et al. 2019; Garaldi et al. 2019; Gaikwad et al. 2021, 2020; Villasenor et al. 2022). This has resulted in the broad picture that the IGM temperature T_0 at mean density rises from about 5,000 K at $z \gtrsim 5.4$, peaks at 15,000 K at $z \sim 3$, and reduces to 7,000 K at $z \lesssim 2$.

Most recent measurements of the statistics of the Lyman-α forest have utilised measurements of a handful of quasars with

* E-mail: nabendu.khan@tifr.res.in

spectrographs on 8m-class telescopes, such as Keck/HIRES, VLT/UVES, and VLT/XSHOOTER (Walther et al. 2018; Boera et al. 2019; Karaçaylı et al. 2022). These measurements have signal-to-noise ratios of 10–30. The typical resolution is medium-to-high, with some spectra having a resolution of $R \sim 4000$–7000, and some $R > 40000$. This allows a measurement of the one-dimensional Lyman-α flux power spectrum to well above $k = 0.1$ s/km, although at these small scales, contamination due to metal-line absorption could become a serious problem (Walther et al. 2018). The resultant precision on the measurement of the flux power spectrum at $k \sim 0.1$ s/km is about 10%. At scales $k \lesssim 0.1$ s/km, the systematic errors are estimated to be about 20% of the statistical errors (Karaçaylı et al. 2022). Low-resolution quasar spectra from surveys like SDSS, BOSS, and eBOSS, containing tens to hundreds of thousands of samples, are also utilized for Lyman-α forest studies. These spectra have limited resolution and signal-to-noise ratios. They are valuable for measuring neutrino masses and background cosmology parameters but cannot effectively constrain warm dark matter cosmologies or the thermal parameters of the IGM due to their insufficient resolution for measuring the Lyman-α forest at small scales with $k > 0.02$ s/km.

The quality of Lyman-α forest data at small as well as large scales is expected to improve dramatically in the near future. DESI will increase the number of quasar spectra by an order of magnitude (DESI Collaboration et al. 2016), leading to order-of-magnitude improvement in constraints on the slope n_s and the running α_s of the primordial curvature power spectrum, and the sum of neutrino masses, and a factor-of-two improvement in constraints on the effective number of neutrino species. On the small scales, improvement is expected from projects such as WEAVE-QSO (Pieri et al. 2016). If the large-scale measurements have a spectral resolution of $R \sim 2000$ and a signal-to-noise ratio of ~ 5, and the small-scale measurements have $R \sim 40000$ and a signal-to-noise ratio of $\gtrsim 30$, the WEAVE-QSO survey will substantially push the amount of intermediate quality data with $R \sim 20000$ and a signal-to-noise ratio of $\gtrsim 20$. This will provide a means to study large-scale effects such as IGM temperature fluctuations. A third improvement is the discovery of more high-redshift quasars by surveys such as LSST and Euclid. In the last decade, the number of quasars with $z > 6$ grew fivefold thanks to deep wide-field optical-infrared surveys such as CFHQS, VISTA/VIKING, Pan-STARRS1, DECaLS, UKIDDS, and UHS. This pushed the number of high-redshift targets for Lyman-α forest studies to 200. The resultant increase in the number of Lyman-α forest measurements has already allowed qualitatively new deductions about the high-redshift universe (Gaikwad et al. 2020; Bosman et al. 2022). LSST will potentially increase this number to 10,000 (for $i < 26$; LSST Science Collaboration et al. 2009). In the nearer term, Euclid is expected to find about 125 quasars with $7 < z < 8$, with potentially a handful of quasars with even higher redshifts (Euclid Collaboration et al. 2019). Over a longer time scale, 30m-class telescopes will give a further boost to Lyman-α forest data, increasing the quality and quantity of spectra to enable new ways of probing the IGM (Skidmore et al. 2015; Maiolino et al. 2013).

These forthcoming improvements to data make it imperative that theoretical modelling of the Lyman-α forest matches measurements in accuracy. Not only would such models be re-quired to be free of numerical errors to a high degree, but they will also need to avoid poorly understood physical assumptions. A large amount of recent work has focussed on this goal. Molaro et al. (2022) and Wu et al. (2019) investigated the effect of ignoring reionization-induced large-scale spatial fluctuations in the IGM temperature on the one-dimensional flux power spectrum of the Lyman-α forest. Molaro et al. (2022) found that for the 10% uncertainties on current measurements, temperature fluctuations do not introduce any biases in the inference of dark matter or thermal parameters. However, for 5% uncertainties, such as those expected in future data, they find systematic shifts of about 1σ. Chabanier et al. (2022) investigated the difference between results from smoothed particle hydrodynamics (SPH) and Eulerian hydrodynamical computation. They found the one-dimensional power spectrum results from both techniques agreed to be better than a percent. They also found that SPH leads to systematic biases in several Lyman-α forest statistics at high redshift because of poor sampling in low-density regions. Bolton & Becker (2009) investigated the effect of mass resolution and box size of SPH simulation on the Lyman-α forest spectra for $2 < z < 5$. They found that the requirements on simulations become more stringent towards higher redshifts because the typical densities probed by the Lyman-α absorption lines are relatively lower. Walther et al. (2021) did a similar analysis for Eulerian codes, finding that a box size of 120 Mpc and a resolution of 30 kpc is required for one-dimensional flux power spectrum predictions that are accurate to 1%. These authors also investigated the accuracy of the interpolation required to span the parameter space to accurately infer model parameters from the data. Bird et al. (2013) also compared Eulerian and SPH methods for modelling the Lyman-α forest. Although they found some differences between the two models at large neutral-hydrogen column densities, there were no significant differences in the column densities probed by the Lyman-α forest. Bird et al. (2020) showed that most simulations of the Lyman-α forest produce a large error in the linear-theory predictions for perturbations in dark matter and baryon densities. They argued that this affects the one-dimensional flux power spectrum predictions at the 5% level. Villasenor et al. (2021) investigated the role of photoheating and photoionization models in Lyman-α forest simulations using their Eulerian code.

In this paper, we turn our attention to the accuracy of the initial conditions used in cosmological hydrodynamical simulations of the Lyman-α forest. In recent times, several different algorithms and codes for generating initial conditions have been used, without a detailed understanding of the robustness of each of them for Lyman-α forest statistics. We set up identical simulations using multiple codes and compare multiple forest statistics. The outline of this paper is as follows. Section 2 describes the current landscape of methods used to set up initial conditions. We describe our set-up in Section 3. Our results are presented in Section 4. We end with a discussion of our findings in Section 5 and a summary in Section 6. Our ΛCDM cosmological parameters are $\Omega_b = 0.0482$, $\Omega_m = 0.308$, $\Omega_\Lambda = 0.692$, $h = 0.678$, $n_s = 0.961$, $\sigma_8 = 0.829$, and $Y_{He} = 0.24$ (Planck Collaboration XVI 2014).

2 INITIAL CONDITIONS IN SPH-BASED COSMOLOGICAL SIMULATIONS

The task of setting up initial conditions for cosmological simulations is to obtain a particle distribution inside a finite volume that represents a density field with Gaussian random fluctuations around the mean. Such a density field should model cosmological density perturbations at sufficiently early times in the matter-dominated era. The resultant density contrast

$$\delta(\mathbf{x}, t) = \frac{\rho(\mathbf{x}, t)}{\bar{\rho}(t)} - 1, \tag{1}$$

can be expanded into its Fourier components as

$$\delta(\mathbf{x}, t) = \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} \mathrm{d}^3 k \, \tilde{\delta}(\mathbf{k}, t) \exp\left(i\mathbf{k} \cdot \mathbf{x}\right). \tag{2}$$

As δ is a Gaussian random field with zero mean, it is fully specified by its power spectrum, $P(k, t)$, defined by

$$\langle \tilde{\delta}(\mathbf{k}, t) \, \tilde{\delta}^*(\mathbf{k}', t) \rangle = P(k, t) \, \delta_{\mathrm{D}}(\mathbf{k} - \mathbf{k}'), \tag{3}$$

where the angle brackets denote an ensemble average, k is the magnitude of the vector $\mathbf{k}$, and δ_{D} is the Dirac delta function. The power spectrum $P(k)$ has units of volume, often chosen to be $(\mathrm{cMpc}/h)^3$. When perturbations are small, the power spectrum is described by a transfer function $T(k, t)$, so that

$$P(k, t) = T^2(k) D^2(t) P_{\mathcal{R}}(k), \tag{4}$$

where $D(t)$ is the linear growth factor, and $P_{\mathcal{R}}(k)$ is the dimensionless primordial curvature power spectrum, taken to be of the form

$$P_{\mathcal{R}}(k) = A_{\mathrm{s}} \left(\frac{k}{k_0}\right)^{n_{\mathrm{s}}}, \tag{5}$$

where, with $k_0 = 0.05 \ \mathrm{cMpc}^{-1}$, measurements favour the power-law index $n_{\mathrm{s}} = 0.961$ and $A_{\mathrm{s}} = 2.2 \times 10^{-9}$ (Planck Collaboration XVI 2014).

In practice, one has to work with quantities defined on a set of discrete points. Consider a Cartesian grid $\mathbf{x}_p = p \Delta x$, with a uniform spacing Δx in the three directions, spanning the simulation volume. We assume that there are n points in each direction, for a total of n^3 points. When the density contrast is evaluated at these discrete points, we can compute the discrete Fourier transform as

$$\tilde{\delta}_{\mathbf{k}_q} = \frac{1}{n^{3/2}} \sum_{p=0}^{n-1} \delta(\mathbf{x}_p) \exp\left(-i\mathbf{k}_q \cdot \mathbf{x}_p\right), \tag{6}$$

where $\mathbf{k}_q = 2\pi q/(n \Delta x)$ for $q = 0, \ldots, n - 1$. The quantity $\tilde{\delta}_{\mathbf{k}_q}$ is related to the quantity $\tilde{\delta}(\mathbf{k})$ defined in Equation (2) by

$$\tilde{\delta}(\mathbf{k}_q) = (\Delta x)^3 \left(\frac{n}{2\pi}\right)^{3/2} \tilde{\delta}_{\mathbf{k}_q}. \tag{7}$$

This allows us to compute the power spectrum from Equation (3) at discrete k_q as

$$P(k_q) = (\Delta x)^3 \langle |\tilde{\delta}_{\mathbf{k}_q}|^2 \rangle, \tag{8}$$

where the angle brackets now denote the mean over all grid points. All three-dimensional power spectra in this paper are computed using Equation (8).

To specify the initial conditions of a cosmological simulation, one must compute the transfer function $T(k)$, defined by Equation (4). This can be done to 0.1% accuracy using Boltzmann solver codes such as CAMB (Lewis et al. 2000) or CLASS (Lesgourgues 2011).

2.1 Generating initial conditions

With the above notation in place, we can now describe the general procedure of setting up initial conditions for a cosmological simulation. This task can be split into two steps. The first step involves creating a uniform distribution of particles. The second step is to displace these particles to achieve a Gaussian random density distribution that has the required power spectrum.

It is not trivial to set up a uniform mass distribution with a finite number of particles (Bertschinger 1998). Simply distributing n particles randomly in the box leads to a density distribution with a non-zero white-noise power spectrum that can lead to non-trivial structure formation even in the absence of additional cosmological power. A conventional solution to this problem is to position the particles in a regular Cartesian cubic grid. This method is widely used in the literature, although it does suffer from problems. For instance, the grid spacing imposes a characteristic length scale in the simulation. The grid also breaks isotropy at small scales by defining preferred directions. These problems are particularly important to address in cosmological models with little or no power at small scales, such as warm-dark-matter cosmologies. An alternative to using a grid distribution is to create a so-called glass-like distribution of particles (Baugh et al. 1995; White 1996). To construct a glass-like distribution, one starts from a random distribution of particle positions, which is then evolved using a repulsive gravitational force. The glass then shows no visible order or anisotropy on scales beyond a few interparticle separations.

Once a uniform particle distribution is obtained, initial conditions for cosmological simulations are derived by shifting these particle positions by displacements designed to get a Gaussian random density distribution with the desired power spectrum. This is achieved in several stages. First, random numbers representing Gaussian random density perturbations with a flat, white-noise power spectrum are sampled using a random number generator, in real or Fourier space, and then transformed so that they have the desired power spectrum (Peacock & Heavens 1985; Bardeen et al. 1986; Salmon 1996). Next, particle displacements are computed from these sampled overdensities. This can be done by mapping the overdensities to gravitational potential perturbations, which then yield the concomitant particle displacements. The displacements are used to modify the uniform particle positions of the initial glass or grid. Lastly, consistent velocities are assigned to the particles.

While this outline of the procedure to generate initial conditions for cosmological simulations is generally uniform across codes, many algorithmic variations exist. Some codes use particle grids while others use particle glasses. Some codes stagger gas and dark matter particles by adding a small initial separation between particle positions, but leave this separation to the user's choice. Some codes sample white-noise density perturbations in real space while others sample them in Fourier space. The computation of particle displacements from the density perturbations is done with a variety of assumptions and accuracy levels of perturbation theory by different codes. Our purpose in this paper is to understand how these differences in initial condition codes affect cosmological simulation results relevant to the Lyman-α forest. In order to do this, we now describe the methods behind the initial

Code	Perturbation sampling	Displacement computation	Velocity computation
NGenIC	Fourier space	Zel'dovich	Equations (10) and (11)
2LPTIC	Fourier space	2LPT	Equation (16)
CosmicIC	real space	Zel'dovich	Equations (10) and (11)
MP-GenIC	Fourier space	Zel'dovich	Equations (17) and (18)
MUSIC	real space	Zel'dovich or 2LPT	Equations (10) and (11)

Table 1. An overview of important methodological similarities and differences between the initial conditions codes investigated in this paper. Although NGenIC, CosmicIC, and MUSIC use the same Equations (10) and (11) for computing velocities, there are small differences in their implementation. NGenIC approximates the quantity $F(z)$ by $\Omega_m^{0.6}(z)$. CosmicIC and MUSIC obtain $F(z)$ using linear perturbation theory, however MUSIC includes the contribution of the radiation energy density while doing so, whereas CosmicIC does not.

See text for details.

conditions codes that we compare in this paper, before looking at the results from cosmological simulations using these codes.

2.2 Initial conditions codes

We compare five initial conditions codes in this paper: NGenIC, 2LPTIC, CosmicIC, MP-GenIC, and MUSIC. We begin by giving a brief description of the algorithms of these codes.

2.2.1 NGenIC

NGenIC (Springel et al. 2005; Angulo et al. 2012) allows the use of either glass and grid particle distributions, and uses the Zel'dovich (1970) approximation to compute the displacements. Gaussian density perturbations with the desired power spectrum are sampled in NGenIC in Fourier space. This is done by drawing random samples x and ϕ from uniform distributions between 0 and 1, and 0 and 2π, respectively. Then, $\tilde{\delta}_\mathbf{k} = Re^{i\phi}$ are the required Fourier-space density perturbations with power spectrum $P(k)$ if $R = (\Delta k)^{3/2}\sqrt{-P(k)\log x}$ with $(\Delta k)^3$ being the Fourier-space volume element defined by Equation (8). The Hermitian symmetry condition, $\tilde{\delta}_\mathbf{k}^* = \tilde{\delta}_{-\mathbf{k}}$, is imposed to ensure that the corresponding real-space perturbations are real.

Using the density perturbations, particle displacements $\mathbf{d}_\mathbf{k}$ are computed using the Zel'dovich approximation as (Zel'dovich 1970)

$$\mathbf{d}_\mathbf{k} = \frac{i\mathbf{k}}{k^2}\delta_\mathbf{k}. \tag{9}$$

These displacements are then transformed to the real space. When the initial uniform particle distribution is on a grid, one can simply add the real-space displacements to the particle positions to get the desired particle distribution. This step can be less than straightforward when the initial uniform particle distribution is not on a grid. In this case, NGenIC obtains the displacements at the particle positions by trilinear interpolation on the displacement grid.

Finally, velocities are assigned to the displaced particles by using the equations of motion to write

$$\mathbf{v}(\mathbf{x}) = aH(z)F(z)\mathbf{d}(\mathbf{x}), \tag{10}$$

where $\mathbf{d}(\mathbf{x})$ are the real-space particle displacements, H is the Hubble constant, z is the redshift at which the initial

conditions are desired, and

$$F(z) = -\frac{d\ln D(z)}{d\ln(1+z)}. \tag{11}$$

In NGenIC, $F(z)$ is approximated to be $\Omega_m^{0.6}(z)$, where $\Omega_m = \rho_m(z)/\bar{\rho}(z)$ is the cosmic matter density relative to the mean. This grid-based NGenIC set-up was used in the simulations belonging to the Sherwood (Bolton et al. 2017) and the Sherwood-Relics (Puchwein et al. 2023) suites.

2.2.2 2LPTIC

2LPTIC (Scoccimarro 1998; Crocce et al. 2006; Scoccimarro et al. 2012) is closely related to NGenIC. The two codes follow identical procedures for setting up the initial uniform particle distributions and density perturbations. They differ in the computation of displacements and velocities, however, as 2LPTIC uses second-order Lagrangian perturbation theory instead of the Zel'dovich approximation discussed above. In this way, the particle displacements in Fourier space are given by

$$\mathbf{d}_\mathbf{k} = \mathbf{d}_\mathbf{k}^{(1)} + \mathbf{d}_\mathbf{k}^{(2)}, \tag{12}$$

where $\mathbf{d}_\mathbf{k}^{(1)}$ is the first-order contribution to the displacement, given by the Zel'dovich approximation via Equation (9), and $\mathbf{d}_\mathbf{k}^{(2)}$ is the second-order contribution given by

$$\mathbf{d}_\mathbf{k}^{(2)} = -\frac{3}{7}\frac{\mathbf{k}\tilde{S}_\mathbf{k}}{ik^2}, \tag{13}$$

where $\tilde{S}_\mathbf{k}$ is the discrete Fourier transform of

$$S = \sum_{i>j}\left[d_{i,i}^{(1)}d_{j,j}^{(1)} - (d_{i,j}^{(1)})^2\right], \tag{14}$$

with $d_{i,j}^{(1)}$ being the derivative of the ith component of the first-order real-space displacement vector $\mathbf{d}$ with respect to the jth component of the position vector,

$$d_{i,j}^{(1)} = \frac{\partial d_i^{(1)}}{\partial x_j} = ik_j d_i^{(1)}. \tag{15}$$

The corresponding velocities are given by

$$\mathbf{v}(\mathbf{x}) = aH(z)\Omega_m^{3/5}(z)\mathbf{d}^{(1)}(\mathbf{x}) + 2aH(z)\Omega_m^{4/7}(z)\mathbf{d}^{(2)}(\mathbf{x}). \tag{16}$$

2.2.3 CosmicIC

CosmicIC (Lukić et al. 2015) is a Zel'dovich code that uses a grid for the initial particle distribution. The code does not itself separate gas and dark matter particles, it instead leaves

this task to the gravity and hydrodynamics solver. To obtain the density field, CosmicIC samples density values $\delta(\mathbf{x}_p)$ from a Gaussian distribution with zero mean and unit variance at each point on a grid in the real space. These white-noise perturbations are then discrete-Fourier transformed to Fourier space. The real and imaginary parts of these Fourier-transformed perturbations then follow a Gaussian distribution with zero mean and variance $\sqrt{n/2}$. These perturbations are then multiplied by $(\Delta k)^{3/2}\sqrt{P(k)/n}$ to obtain Gaussian density perturbations in the Fourier space with the given power spectrum $P(k)$. Displacements are then computed from these Fourier space density perturbations using the Zel'dovich approximation of Equation (9). The corresponding velocities are computed using Equations (10) and (11), however unlike NGenIC, CosmicIC does not approximate the quantity $F(z)$ with $\Omega_m^{0.6}(z)$. Instead, CosmicIC computes $F(z)$ using linear perturbation theory. While doing so it ignores the radiation energy density.

2.2.4 MP-GenIC

MP-GenIC (Bird et al. 2020) allows for various ways of initialising the particle positions. Bird et al. (2020) showed that initialising dark matter particles on a grid and gas particles on a glass reproduces the expected linear evolution of the difference between the power spectra of gas and dark matter. This is the strategy we adopt while using MP-GenIC in this paper. MP-GenIC generates Gaussian density perturbations with a given density power spectrum in an identical manner to NGenIC and 2LPTIC. The gas and dark matter densities are then obtained by scaling the total matter density perturbations with the Fourier-space gas and dark matter density fractions obtained from CLASS (Lesgourgues 2011). The corresponding displacements are obtained by means of the Zel'dovich approximation, via Equation (9). This displacement field is then transformed to real space. Gas particle displacements are then obtained using trilinear interpolation. Velocities are obtained from CLASS as transfer functions

$$v_{\rm CDM}(k) = \dot{\delta}_{\rm CDM}(k) = -\frac{\dot{h}(k)}{2}, \tag{17}$$

and

$$v_{\rm b}(k) = \dot{\delta}_{\rm b}(k) = -\theta_{\rm b}(k) - \frac{\dot{h}(k)}{2}, \tag{18}$$

where h is the synchronous gauge density perturbation, and $\theta_{\rm b}$ is the baryon velocity divergence, both computed by CLASS. These velocity transfer functions are sampled similar to the density transfer functions. The resultant Fourier-space velocities are then transformed to real space and interpolated using CIC. MP-GenIC avoids spurious gravitational coupling between gas and dark matter particles by evolving the pre-displaced combined distribution of the two types of particles with reversed gravity for a small number of time steps.

2.2.5 MUSIC

MUSIC (Hahn & Abel 2011) allows initialising the particle load using either a grid or a glass. We use the grid in this work. In contrast with NGenIC, MUSIC obtains real-space Gaussian density perturbations with a given power spectrum
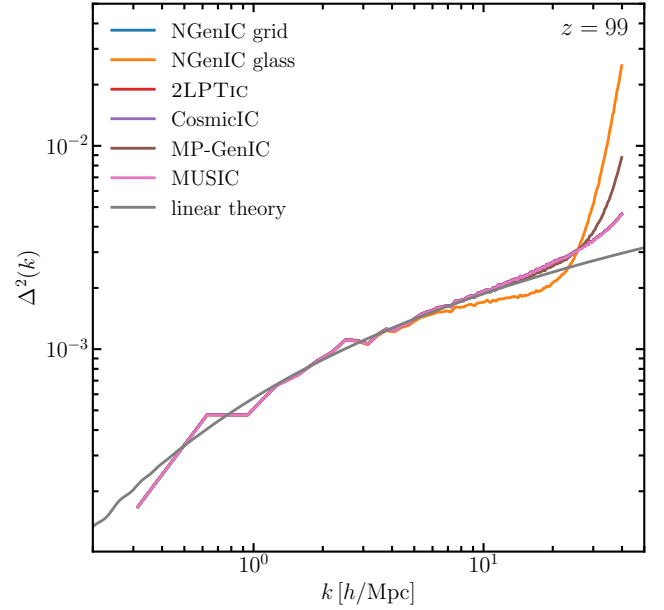

Figure 1. The three-dimensional power spectrum of the total matter density in the initial conditions used for our simulations, at $z = 99$. We show results from the five codes that we study. For NGenIC, we also show the result when using a glass-like instead of a grid-based particle distribution for the initial unperturbed matter density. Most of the curves overlap each other for most of the range of k values. The grey curve is the linear theory power spectrum computed using CLASS (Lesgourgues 2011). We see that in the initial conditions, the matter density power spectrum between the codes agrees very well. The only significant outliers are the glass-based NGenIC and MP-GenIC set-ups.

by sampling a Gaussian variate with zero mean and unit variance on a grid. This Gaussian variate is then convolved with the inverse Fourier transform of $\sqrt{P(k)}$ to get the desired density perturbations $\delta(\mathbf{x})$ on the grid. Particle displacements are then obtained in a similar way as NGenIC following Equation (9). Velocities are also computed similar to NGenIC, using Equations (10) and (11). However, unlike NGenIC, MUSIC computes the quantity $F(z)$ using linear perturbation theory. While doing so, unlike CosmicIC, MUSIC includes the radiation energy density. Finally, corrections are applied to these velocities to compensate for inaccuracies caused by discreteness.

2.3 Initial conditions for different particle initialisation choices

Our fiducial simulations consist of five runs with the five initial condition set-ups from Table 1. Figure 1 shows the three-dimensional spherically-averaged power spectrum of the total matter density contrast at $z = 99$ from our five fiducial simulations. Also shown in this figure is the linear theory result at this redshift, obtained using CLASS (Lesgourgues 2011). The figure also shows the power spectrum from the glass-based NGenIC run, to compare to the corresponding grid-based run.

We see that all grid-based simulations reproduce the theoretical linear power spectrum extremely well. (Most curves

in Figure 1 completely overlap with each other.) The only case with some difference is 2LPTIC, where the use of the second-order perturbation theory results in differences from our fiducial run, but even these differences are sub-percent. In contrast to these results, a large difference is seen in the glass-based NGenIC run, which deviates from its grid-based counterpart with suppression in power of close to 20% at $k = 20\,h/\mathrm{cMpc}$ and a strong enhancement in power of up to a factor of 2 at larger values of k. This excess power in the glass-based set-up is the well-known k^4 enhancement due to the growth of random gravitational instabilities during the creation of the glass configuration (Bird et al. 2020). Indeed, a similar enhancement is also seen in the MP-GenIC case, although this is relatively smaller because MP-GenIC uses glass only for the baryons.

While such small-scale enhancement suggests a numerical error, Bird et al. (2020) pointed out that there might be advantages in initialising gas particles on a glass. They argued that the MP-GenIC set-up, which uses distinct transfer functions for gas and dark matter, and uses a glass-like configuration for gas particles, can reproduce the linear theory evolution of the power difference between gas and dark matter. Figure 2 shows this at $z = 9$. We show the difference in the dark matter and gas power spectra in four runs, by plotting the quantity $\eta = (\tilde{\delta}_{\mathrm{dm}} - \tilde{\delta}_{\mathrm{gas}})/2$ at $z = 9$. (The simulation details on the evolution from $z = 99$ to $z = 9$ are described in the next section.) The green curve shows the power difference from our fiducial grid-based NGenIC run, and the red curve shows the same quantity for our fiducial MP-GenIC run. The yellow curve shows how the result changes when we use identical transfer functions for gas and dark matter in the MP-GenIC run. The blue curve shows the power difference when we use a grid-based initialisation for gas particles in the MP-GenIC run. For all four cases we have shown the difference between simulated and the linear theory value of η. In the linear theory, the power in dark matter and baryons is calculated from the CLASS output using the appropriate columns in the transfer function and total matter power spectrum files. The transfer functions are calculated as $\delta_{\mathrm{cdm/b}} = \delta_{\mathrm{cdm/b}}/\delta_{\mathrm{total}}$, where $\delta_{\mathrm{cdm/b}}$ are from the CLASS outputs and $\delta_{\mathrm{total}} = (\Omega_{\mathrm{b}}\delta_{\mathrm{b}} + \Omega_{\mathrm{cdm}}\delta_{\mathrm{cdm}})/(\Omega_{\mathrm{b}} + \Omega_{\mathrm{cdm}})$. The powers for both species are given by $\delta_{\mathrm{cdm/b}} = \delta_{\mathrm{cdm/b}}\sqrt{P_{\mathrm{tot}}}$ and P_{tot} is obtained from the CLASS power spectrum outputs. As noted by Bird et al. (2020), we see here that only the fiducial MP-GenIC run manages to reproduce the expected linear evolution. Removing the glass, or removing the difference between the gas and dark matter transfer functions results in a significant disagreement with the linear theory.

We will now see how these particle initialisation choices affect the small-scale structure in the Lyman-α forest.

3 NON-LINEAR EVOLUTION

Our procedure for computing the non-linear gravitational and hydrodynamical evolution is identical to the set-up used in the Sherwood Simulation Suite (Bolton et al. 2017). We perform our simulations using the SPH code P-Gadget3, the well-known modified version of the publicly available Gadget2 code (Springel et al. 2005). In this code the gravitational forces are computed using the TreePM algorithm where short-range forces are obtained using a tree method and long-
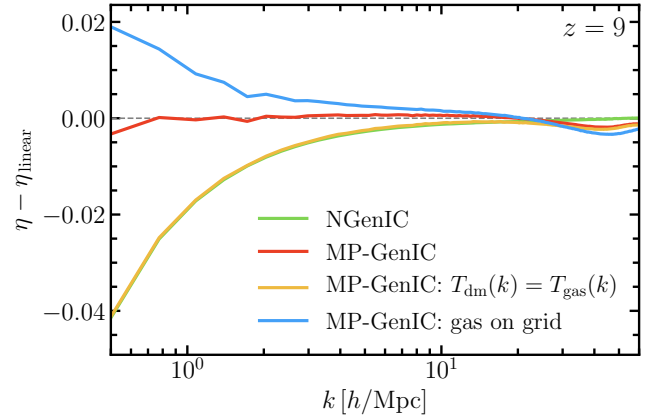

Figure 2. The difference between the linear theory value of $\eta = (\tilde{\delta}_{\mathrm{dm}} - \tilde{\delta}_{\mathrm{gas}})/2$ and its value in four of our simulations at $z = 9$. The green and red curves show the power difference from our fiducial grid-based NGenIC run and our fiducial MP-GenIC run, respectively. The yellow curve shows the result when we use identical transfer functions for gas and dark matter in the MP-GenIC run, while the blue curve shows the same when we use a grid-based initialisation for gas particles in the MP-GenIC run. As previously found by Bird et al. (2020), we see that only fiducial MP-GenIC simulation recovers the correct linear theory evolution.

range forces are computed using mesh-based Fourier methods. In the tree algorithm, particles are grouped together according to their distances and the force from each group is approximated using a multipole expansion. In the particle-mesh (PM) algorithm the density field is realised on a mesh grid using CIC interpolation and the gravitational potential is constructed by solving the Poisson's equation. The mesh dimension in each direction is set by the `PMGRID` parameter, which we set to the cube root of the particle number in all simulations presented in this paper. To avoid a singularity in the force calculation when the particles move close to each other, a softening parameter l_{soft} is introduced. We set this to 1/25th of the mean inter-particle separation. Apart from MP-GenIC, we use P-Gadget3 for splitting particles into gas and dark matter. The dark matter and gas positions are then calculated by adding constant vectors to the total matter positions, keeping the centre of mass at the position of the original particle. The constant separations are

$$\delta_{\mathrm{gas}} = \frac{1}{2}\frac{\Omega_{\mathrm{m}} - \Omega_{\mathrm{b}}}{\Omega_{\mathrm{m}}}d, \quad \text{and}$$

$$\delta_{\mathrm{dm}} = \frac{1}{2}\frac{\Omega_{\mathrm{b}}}{\Omega_{\mathrm{m}}}d,$$

where $d = L/N$ is the inter-particle spacing. Our fiducial simulation, described below, has $d = 78.125\,\mathrm{ckpc}/h$, $\delta_{\mathrm{gas}} \sim 33\,\mathrm{ckpc}/h$ and $\delta_{\mathrm{dm}} \sim 6\,\mathrm{ckpc}/h$. These separations are added to the total matter positions along all three directions, so that $x_{\mathrm{gas}} = x_{\mathrm{gas}} - \delta_{\mathrm{gas}}$, $y_{\mathrm{gas}} = y_{\mathrm{gas}} - \delta_{\mathrm{gas}}$, and $z_{\mathrm{gas}} = z_{\mathrm{gas}} - \delta_{\mathrm{gas}}$ for the gas particles, and $x_{\mathrm{dm}} = x_{\mathrm{dm}} + \delta_{\mathrm{dm}}$, $y_{\mathrm{dm}} = y_{\mathrm{dm}} + \delta_{\mathrm{dm}}$, and $z_{\mathrm{dm}} = z_{\mathrm{dm}} + \delta_{\mathrm{dm}}$ for the dark matter particles. Both gas and dark matter particle velocities are assigned to be the same as that of the original total matter particle. In the case of the MP-GenIC, the particles are separately supplied to P-Gadget3 as part of the initial conditions.

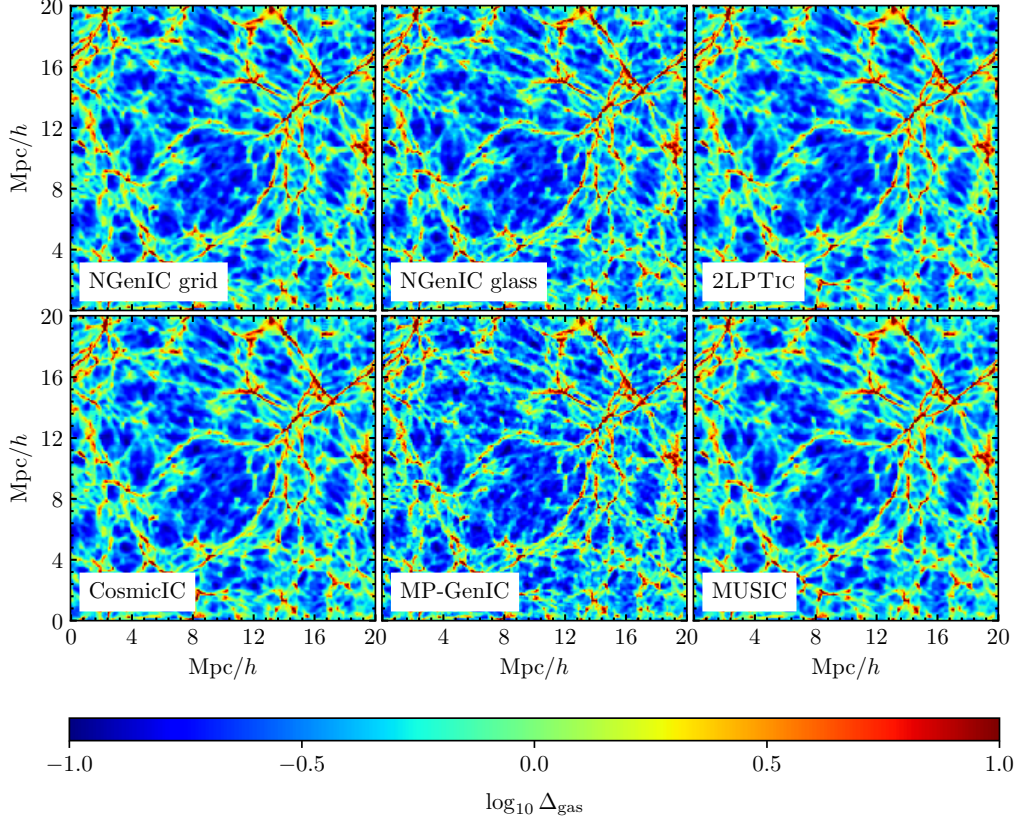We run our simulations down to $z = 2$ and examine results

Figure 3. The gas density distribution at $z = 3$ for our six primary simulations. Each slice shows the gas density projected over 78.125 ckpc/h. The simulations produce similar gas distributions, thanks to our use of the identical random phases in all of them.

at $z = 9, 5, 4, 3,$ and 2. In addition to the cosmological evolution of baryons and dark matter, P-Gadget3 implements photoionization, photoheating, radiative cooling. We use the QUICK_LYALPHA flag in the code to speed up the simulation by making gas particles with overdensities of greater than 10^3 with temperature less than 10^5 K collisionless and removing them from the hydrodynamic calculation. This does not affect our study of the low-density IGM. The metagalactic UV background is modelled according to a modified version of the Haardt & Madau (2012) reionization model (see Bolton et al. 2017 for details).

Our fiducial simulations use $N = 2 \times 256^3$ particles to present the gas and the dark matter in a computational box of comoving length $20\,h^{-1}$Mpc. We refer to these as the 20–256 runs. Periodic boundary conditions are implemented. The mass of dark matter particle is $M_{\rm dm} = 3.44 \times 10^7\,{\rm M}_\odot/h$ and the gas particles are of mass $M_{\rm gas} = 6.38 \times 10^6\,{\rm M}_\odot/h$. We also study higher resolution 20–512 and 20–1024 simulations. As their names suggest, these simulations use 2×512^3 and 2×1024^3 particles in the same volume as the 20–256 simulations. In the 20–512 simulations, the dark matter and gas masses are $M_{\rm dm} = 1.72 \times 10^7\,{\rm M}_\odot/h$, and $M_{\rm gas} = 3.19 \times 10^6\,{\rm M}_\odot/h$, and we use PMGRID $= 512$ and $l_{\rm soft} = 1.5625$ kpc. For the 20–1024 simulations, these quantities take the values $M_{\rm dm} = 0.86 \times 10^7\,{\rm M}_\odot/h$, $M_{\rm gas} = 1.595 \times 10^6\,{\rm M}_\odot/h$, PMGRID $= 1024$ and $l_{\rm soft} = 0.78125$ kpc.

In order to compute the dark matter and gas density power

spectra, we project the output of the hydrodynamical simulations on a uniform Cartesian grid. The densities for the dark matter are calculated on a regular grid from the particle positions using the CIC kernel (Hockney & Eastwood 1981). The gas particles are projected on the grids by integrating the SPH kernel on grid cells. The density contrast $\delta = \rho/\bar{\rho} - 1$ is calculated for both dark matter and gas particles on each grid. Then the power spectrum is then calculated using Equation (8). We set the number of cells in each direction for the grid in k-space to the cube root of the number of particles. While computing the power spectrum of the dark matter density, we deconvolve the CIC kernel.

The Lyman-α forest sightlines are extracted from the simulation box by selecting random points on the xy plane, and choosing all lines to go in the z direction. Then the Lyα optical depth is calculated as

$$\tau(v) = \frac{3 A_{2p1s} \lambda_{\rm Ly\alpha}^3}{8\pi H(z)} \int_{-\infty}^{\infty} n_{\rm HI}\phi(v){\rm d}v, \qquad (19)$$

where $n_{\rm HI}$ is the neutral hydrogen number density, $\lambda_{\rm Ly\alpha}$ is the rest-frame Lyα wavelength, A_{2p1s} is the Einstein A coefficient and $\phi(v)$ is the Voigt line profile. The Lyα transmission is then $F = \exp(-\tau)$ Following standard practice for such optically thin simulations, we normalise the transmission such that the mean transmission has the observed value.

While deriving the one-dimensional power spectrum of the Lyman-α transmission, we define the Lyman-α flux contrast
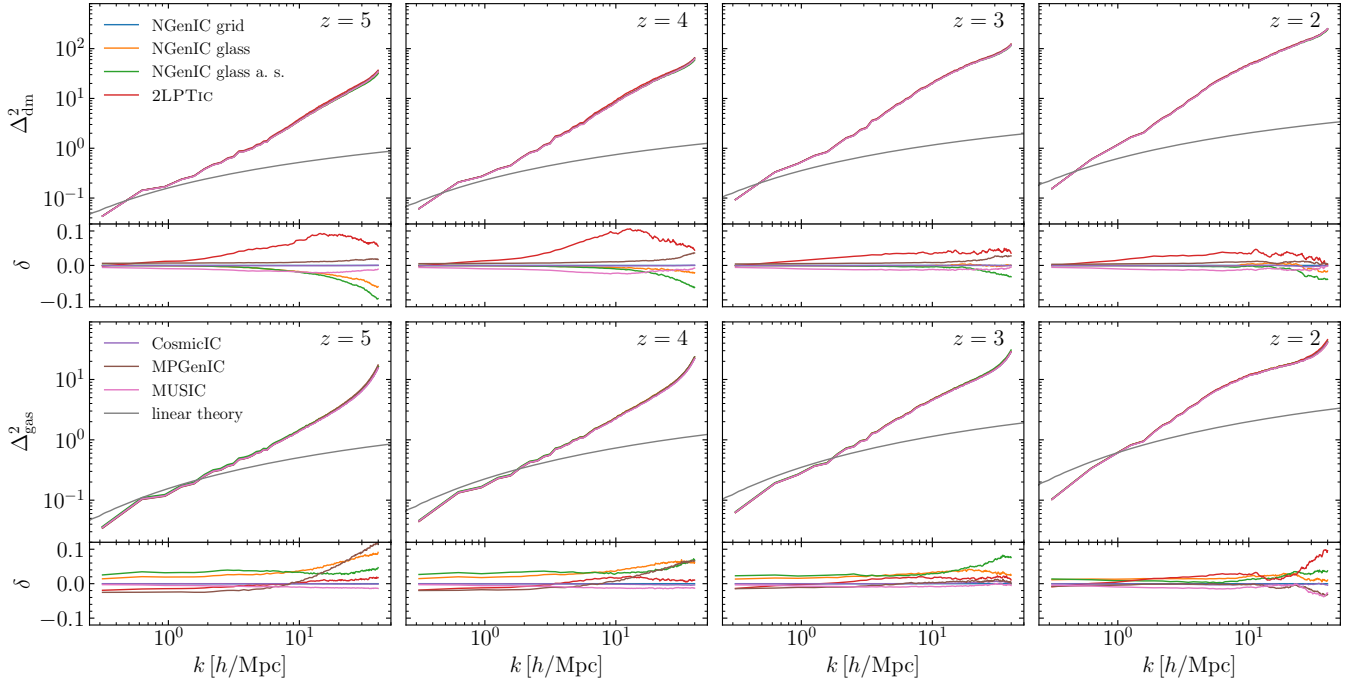
Figure 4. The three-dimensional dimensionless power spectrum of the dark matter (top row) and gas density (bottom row) at $z = 2, 3, 4$, and 5. Each panel shows results from our five primary simulations. Also shown are results from a glass-based NGenIC run, and a glass-based NGenIC run that uses adaptive softening (indicated by 'a. s.' in the legend). The linear theory power spectrum is also shown at each redshift. The difference panels show the relative difference from the grid-based NGenIC results.

analogous to the density contrast, given by $\delta_F = F/\bar{F} - 1$, where $\bar{F}$ is the mean flux. The power pectrum is then computed as

$$P_{F,1D}(|k|) = v_{\text{scale}} \langle \tilde{\delta}_F \tilde{\delta}_F^* \rangle, \tag{20}$$

where $v_{\text{scale}} = aH(z)L$, with L being the length of the box, $\tilde{\delta}_F$ is the Fourier transform of δ_F, and the ensemble average is computed by averaging $\tilde{\delta}_F \tilde{\delta}_F^*$ over 5,000 sightlines for each $|k|$.

We also perform a set of simulations with adaptive softening of gas particles. This means that the gravitational softening length of gas particles is set by the SPH smoothing length. As mentioned above, the default gravitational softening length for gas particles in our fiducials simulations is the same as that of dark matter particles, and is set to 1/25th of the mean interparticle separation (Bolton et al. 2017). Adaptive softening can be enabled in P-Gadget3 by setting the `ADAPTIVE_GRAVSOFT_FORGAS` and `ADAPTIVE_GRAVSOFT_FORGAS_HSML` flags. The SPH smoothing kernel length is proportional to the comoving distance to the 32nd neighbour (Springel et al. 2001). As a result, at high redshifts, when the universe is more homogeneous, the softening length can be as large as three times the mean interparticle separation. As we will see below, this suppresses small-scale power in the gas density distribution.

4   RESULTS

Our fiducial simulations consist of five runs with the five initial condition set-ups from Table 1. The simulation specifications are as described above. We now present the results from

these simulations. We will also occasionally refer to variations on these simulations. There are four such variations that we consider. One of these uses NGenIC with a glass-like initial particle load. Another employs the same glass-based NGenIC set-up but with adaptive softening. In all other respects, these runs are identical to our fiducial 20–256 runs. Our third and fourth variations are the 20–512 and 20–1024 counterparts of our fiducial 20–256 runs. These high-resolution runs have been mentioned above in Section 3.

For each of the six simulations shown in Figure 1, two-dimensional gas density distribution slices at $z = 3$ are shown in Figure 3. Each slice has the thickness of a single grid cell, which in our case is 78.125 kpc/h. The familiar non-linear large-scale structure at this redshift is clearly seen. Furthermore, the structure in all simulations appears to be visually identical, showing that we are using identical random numbers while generating the initial conditions.

4.1   Nonlinear matter power spectrum

The top four panels of Figure 4 show the three-dimensional power spectrum of the dark matter density contrast at redshifts 2, 3, 4 and 5, in our five fiducial simulations. We also show results from the simulation with glass-based NGenIC and the one with adaptive softening. The linearly extrapolated matter power spectrum at respective redshifts is also shown. At the bottom of each panel the difference relative to the grid-based NGenIC run is shown. Compared to Figure 1, we can now see the power enhancement relative to linear theory at small scales due to nonlinear evolution. As anticipated in Figure 3, the power spectra agree between the different
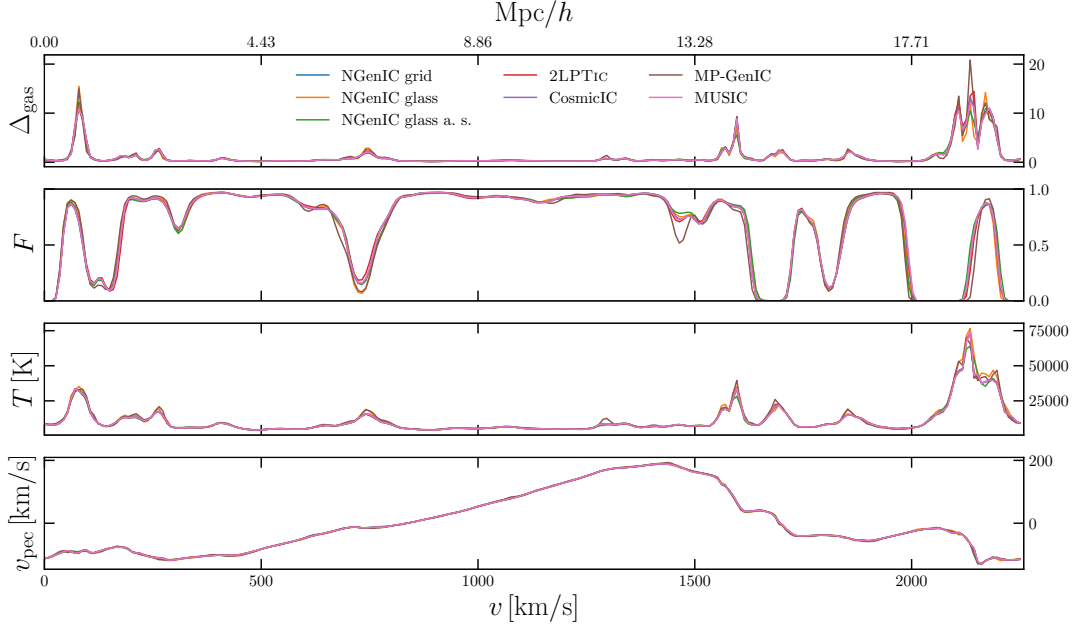
Figure 5. The gas overdensity, Lyman-α transmitted flux, gas temperature and gas peculiar velocity along a randomly chosen one-dimensional skewer from our simulation boxes. Each panel shows results from our five primary simulations. Also shown are results from a glass-based NGenIC run, and a glass-based NGenIC run that uses adaptive softening (indicated by 'a. s.' in the legend).

simulations to a large degree, but interesting differences are now visible.

Overall, all power spectra are consistent within 10% over the scales shown in Figure 4. The largest deviation from the fiducial run is seen in the 2LPTIC case, which exhibits more power relative to the NGenIC run. The difference is at most 10% at the scale of $k = 10 \, h/\text{cMpc}$, and decays with time. The dark matter distribution in our glass-based NGenIC run has less dark matter power at small scales relative to the NGenIC grid. These differences appear to be enhanced when we use adaptive softening, but in either case the differences decay with time. Our CosmicIC results are nearly identical to those from the NGenIC run. The MP-GenIC and MUSIC cases show only small differences that appear well below 5% at all redshifts shown here. The difference between the fiducial run and the MUSIC case can potentially be partly explained by the small difference between the value of the growth factor used in the two codes (see Section 2).

The bottom four panels of Figure 4 show the three-dimensional gas density power spectrum from the simulations shown in the top panel, along with the linear theory prediction, at redshifts 2, 3, 4 and 5. As before, we show the relative difference from the grid-based NGenIC run at the bottom of each panel. The simulations agree with each other reasonably well, with differences of at most 10% in the gas density power spectrum. The greatest deviation is seen in the glass-based runs, NGenIC and MP-GenIC. Both of these runs produce gas power larger by $\sim 10\%$ at small scales relative to the NGenIC using a grid. The differences appear to increase towards the small scales resolved by the simulation. This power enhancement is related to that seen in the initial conditions, shown in Figure 1, where we understood it to be the result of the nonlinearities in the process of creating the glass. We also see that this power enhancement reduces

with time as we go from $z = 5$ to 2 in Figure 4. Further, the use of adaptive softening seems to remove the power excess, as noted previously by O'Leary & McQuinn (2012). If we compare the dark matter and gas density power spectra in Figure 4, we can see that dark matter has more power than gas at all scales and all redshifts. This can be understood because of the exclusion of the gas particles that were turned into stars in biased high-density regions. This results in the attenuation of power at small as well as large scales for the baryons (Kulkarni et al. 2015).

4.2 One-dimensional Lyman-α transmission power spectrum

We now turn to the Lyman-α transmission properties of our simulations. Figure 5 shows one-dimensional sightlines, 'skewers', from the simulations shown in Figure 4. All sight-lines are shown at $z = 3$. We show the gas density, the Lyman-α flux, gas temperature, and gas peculiar velocity along a randomly chosen sightline. We normalise the simulations to identical mean transmission values. We choose these to be the measurements by Becker et al. (2013). As before, while the seven simulations appear very similar, there are small differences. These appear to be largest near the density peaks and troughs, with the MP-GenIC run being the widest outlier. The Lyman-α transmission exaggerates these differences, so that there is up to a 30% difference in the transmission. Small differences also appear to be present in the gas temperature. Most importantly, there are small differences in the peculiar velocity. These small peculiar velocity differences will, however, have a large effect on the small-scale power spectrum of the Lyman-α forest.

Figure 6 shows the one-dimensional Lyman-α flux power spectrum from all the seven simulations at redshifts 2, 3,
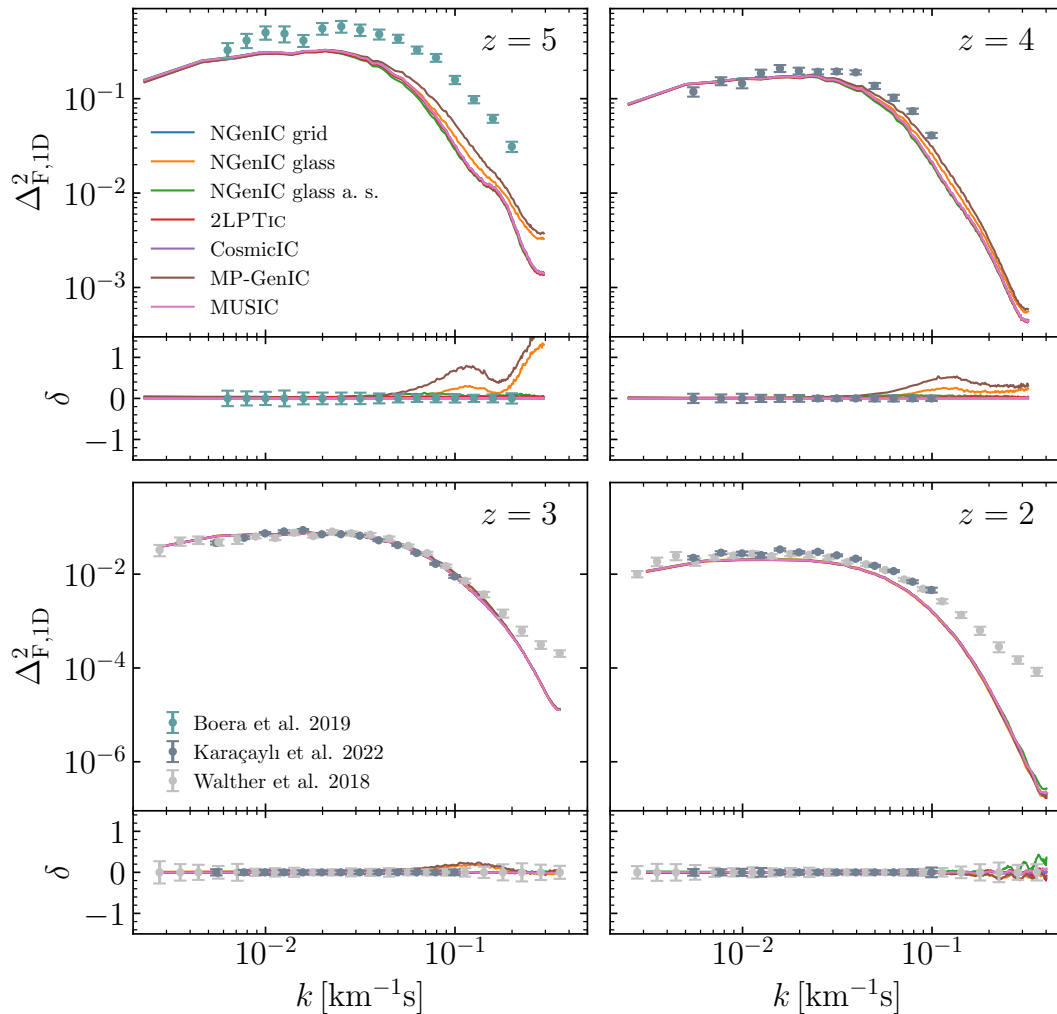
Figure 6. The one-dimensional Lyman-α transmission power spectrum at $z = 2, 3, 4,$ and 5. Each panel shows results from our five primary simulations. Also shown are results from a glass-based NGenIC run (orange curve), and a glass-based NGenIC run that uses adaptive softening (green curve, indicated by 'a. s.' in the legend). We also show measurements by Walther et al. (2018), Boera et al. (2019), and Karaçaylı et al. (2022). These data are only shown for context and as a measure of the current uncertainties. We make no attempt at fitting the data with our models, which explains the mismatch between the simulations and the observations at some redshifts and scales. The error panels show differences relative to the result from the grid-based NGenIC simulation. All spectra are normalised to the observed mean transmission from Becker et al. (2013).

4 and 5. We show the dimensionless Lyman-α transmission power spectrum. At the bottom of each panel, we show differences relative to the grid-based NGenIC run. The figure also shows measurements by Karaçaylı et al. (2022), Walther et al. (2018), and Boera et al. (2019). While we make no attempt the fit our simulations to these data, the uncertainty estimates on the measurements provide a context while examining the differences in our simulations. A better agreement would require a well-calibrated thermal model run at higher resolution. We now see that the glass-based runs using NGenIC and MP-GenIC severely disagree with the grid-based result. The differences are worse in the case of MP-GenIC. At $k = 0.1$ s/km at $z = 5$, MP-GenIC can yield difference in excess of 50%. This is not only far in excess of the data uncertainty, but is also enough to significantly bias inferences of parameters such as the warm dark matter particle mass and the IGM temperature. We will investigate the causes behind

these differences below. The differences decay with time to reduce to percent levels at $z = 2$. It is interesting to note that the result from the glass-based NGenIC model deviates less from the grid-based result than MP-GenIC. But both MP-GenIC and the glass-based NGenIC runs show very similar k-dependence of the power spectrum, indicating a common origin for their behaviour. We also see that introducing adaptive softening in the glass-based NGenIC run removes the differences from its grid-based counterpart. All other runs are in excellent agreement with each other.

5 UNDERSTANDING THE SMALL-SCALE DIFFERENCES

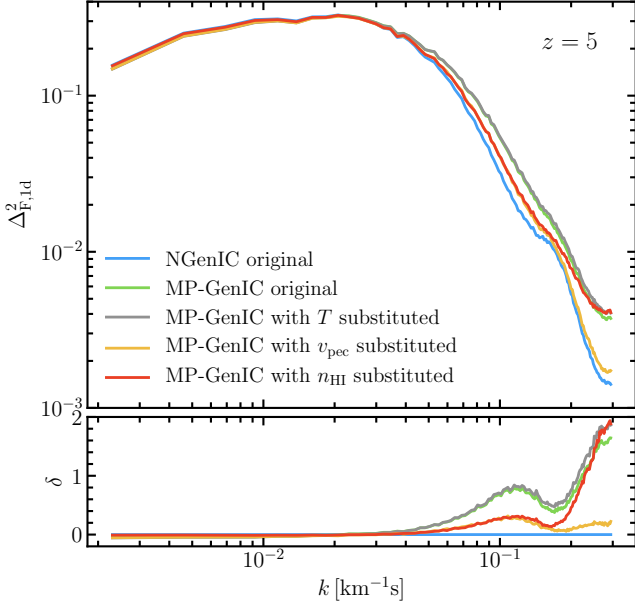Figure 7 helps us understand the reasons behind the large difference that we see between the Lyman-α flux power spec-

Figure 7. The one-dimensional Lyman-α transmission power spectrum from our NGenIC and MP-GenIC runs at $z = 5$. The blue and green curves show our fiducial grid-based NGenIC and MP-GenIC runs, respectively. The other curves show the power spectrum from the MP-GenIC run when the temperatures (grey curve), peculiar velocities (yellow curve), or gas densities (red curve) in this simulation are substituted by their corresponding values from the NGenIC run. The simulations are normalised to the observed mean flux in all cases, as described in the text. The bottom panel shows the fractional change with respect to the NGenIC result.

trum from our NGenIC and MP-GenIC runs. This figure shows the Lyman-α flux power spectrum at $z = 5$. The green and blue curves show our fiducial results from MP-GenIC and NGenIC, respectively. These power spectra are the same as those in Figure 6 above. The other curves show the power spectrum from the MP-GenIC run when the temperatures (grey curve), peculiar velocities (yellow curve), or gas densities (red curve) in this simulation are substituted by their corresponding values from the NGenIC run. The simulations are normalised to the observed value of the mean flux in all cases. The bottom panel shows the fractional change with respect to the NGenIC result. We see that there is no significant difference between the gas temperatures in the two simulations. When the gas temperature in the MP-GenIC run is substituted by the NGenIC run, the Lyman-α flux PS shows almost no change. But the effect is significant when we substitute the gas velocity or the gas density. Gas velocity is the greater effect between the two. At $k > 0.2$ s/km, almost the entire difference between the power spectra comes from gas velocity.

As Figure 8 shows this behaviour is entirely due to the numerical choice of using a glass-like configuration for the gas particles in our MP-GenIC run. The figure shows the relative change in the dark matter power spectrum (top panel), the gas density power spectrum (middle panel), and the one-dimensional Lyman-α flux power spectrum. Recall that there are two important differences between our NGenIC and MP-GenIC simulations: (*a*) the gas and dark matter initial power
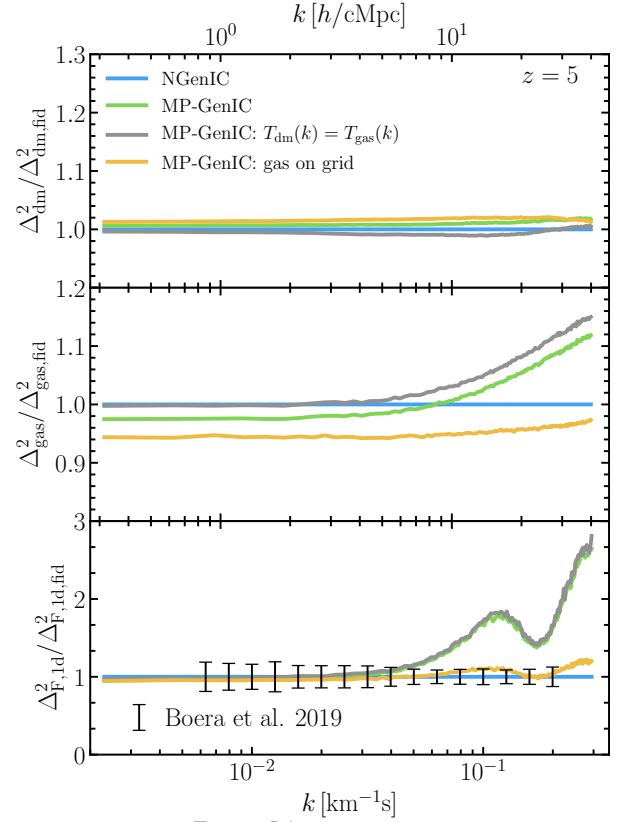

Figure 8. Effect of modifying the gas particle set-up in our MP-GenIC run on the dark matter density power spectrum (top panel), gas density power spectrum (middle panel), and the one-dimensional Lyman-α flux power spectrum. The blue and green curves show our fiducial grid-based NGenIC and MP-GenIC runs, respectively. The grey curves show results from an MP-GenIC run in which identical transfer functions are used for gas and dark matter density perturbations. The yellow curves show results from an MP-GenIC run with different transfer functions but with the unperturbed initial positions of gas particles as well as dark matter particles following a uniform grid. The error bars in the bottom panel are uncertainties on the power spectrum measurements by Boera et al. (2019).

spectra are identical in the NGenIC run, but these are different in the MP-GenIC run, and (*b*) gas particles are initialised on a grid in the NGenIC run whereas they are initialised on a glass in the MP-GenIC run. Figure 8 investigates what happens when we remove each of these differences between the two runs one at a time. The blue and green curves in this figure show various power spectra from our fiducial NGenIC and MP-GenIC runs. The grey curve shows power spectra from an MP-GenIC run in which we use identical initial transfer functions for gas and dark matter. The yellow curves show power spectra for an MP-GenIC run in which the gas and dark matter transfer functions are distinct as before but now the gas particle positions are initialised on a grid. We see that the dark matter power spectra are nearly identical in all cases. The small differences of $\lesssim 2\%$ result partly from the choice between the use of the total matter power spectrum and dark matter power spectrum, but also due to a backreaction of the large differences in the gas power spectra. However, the gas
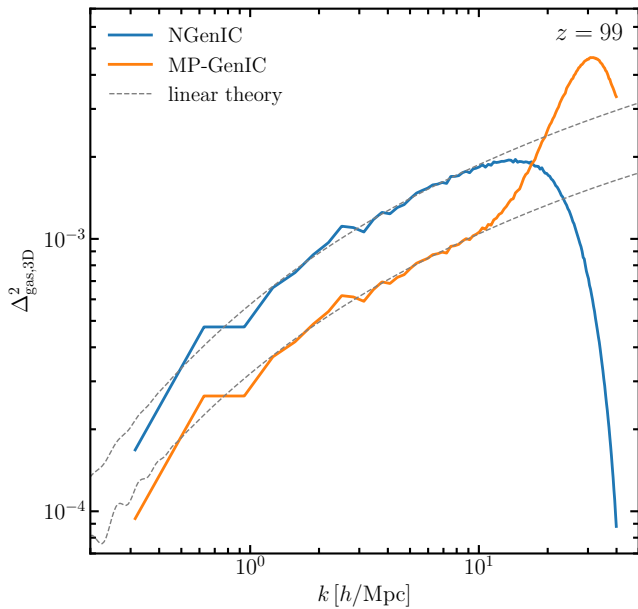
Figure 9. The three-dimensional power spectrum of the gas density at $z = 99$ in our NGenIC and MP-GenIC simulations. The difference in the overall amplitude is due to the difference in the transfer functions used in the two runs. The differences at the small scales are due to the use of glass in MP-GenIC. The dashed curves show the linearly extrapolated power spectra in the two cases.

Figure 10. The top panel shows how the Lyman-α flux power spectrum ratio between NGenIC and MP-GenIC changes with resolution. The middle panel shows the change in the MP-GenIC run with resolution. The bottom panel shows the same for our NGenIC run. Even at the highest resolution, there is a $\sim 50\%$ difference in the Lyman-α flux power spectrum from the two runs. The error bars in each panel are uncertainties on the power spectrum measurements by Boera et al. (2019). These are shown simply for a comparison with the differences between the simulation curves.

density power spectra are significantly different. Here, we first see that there is a broadband difference between the NGenIC and MP-GenIC runs because of the difference in the transfer functions. This difference vanishes at large scales when we use the same transfer functions for gas and dark matter in the MP-GenIC run. Next, at small scales, the order 10% differences in the gas density power spectra are due to the adoption of glass-like initial conditions. We see that this difference goes away as we shift from glass to grid. Figure 9 shows that these differences in the gas power spectra are already present at $z = 99$. (In Figure 9, we also show the linearly extrapolated power spectra for the two simulations. We see that both simulations show deviation from the linear theory prediction at small scales. The NGenIC run underpredicts the small-scale power. This is the result of smoothing due to the SPH kernel. The MP-GenIC run overpredicts the small-scale power. This is because of the use of glass.) Finally, turning to the Lyman-α flux, the bottom panel of Figure 8 shows that the Lyman-α flux power spectrum does not change at all with the change in the gas initial power spectrum. However, it shows a big change when we shift from glass to grid. Indeed, when we used a grid in our MP-GenIC run, the flux PS is nearly identical to the NGenIC result. We are thus in a situation that the set-up designed to reproduce the correct linear-theory behaviour ends up introducing significant changes in the Lyman-α transmission power spectra.

All of the above results used our default 20–256 runs. However, this spatial resolution has been known to be insufficient for convergence for the Lyman-α flux PS at $z = 5$ (Bolton & Becker 2009; Rogers & Peiris 2021; Doughty et al. 2023). In Figure 10, we look into the effect on the flux power spectrum of the spatial resolution of our runs. We now do our NGenIC
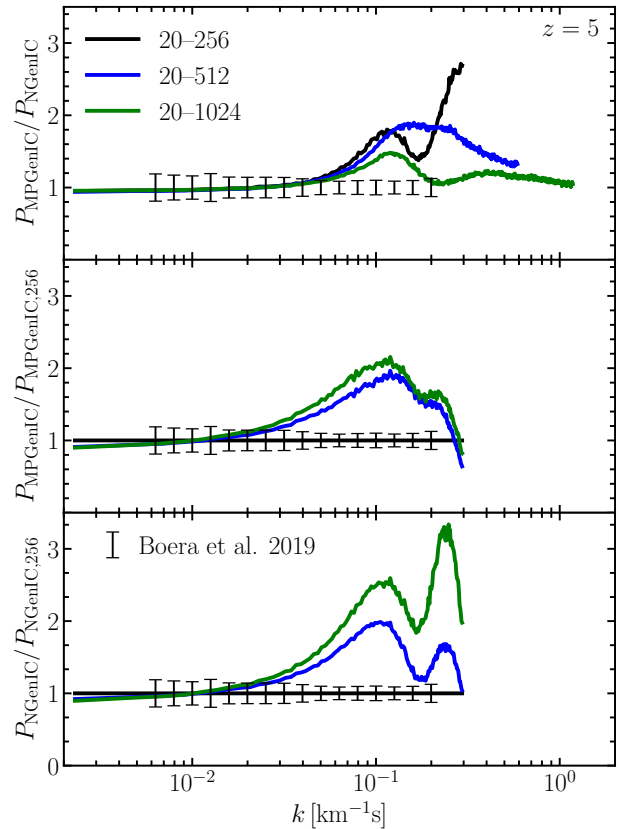
and MP-GenIC runs at 20–512 and 20–1024 resolutions. In each case, we set the value of `PMGRID` to the cube root of the particle number, and we use a softening length that is $1/25$ of the mean inter-particle spacing. See Appendix D for how these results change if we change these settings. Note that 20–1024 is the resolution used by many Lyman-α forest models in the literature. We see in Figure 10 that even at our highest resolution there is upto 50% difference in the flux PS at $z = 0.1$ s/km between our NGenIC run and MP-GenIC runs. It is also important to note that the difference is scale dependent, thus precluding the possibility of applying simple scale-independent 'resolution corrections' to low-resolution runs. (The enhancement in power at $k \sim 0.25$ s/km, is similar in shape to the enhancement that Iršič et al. (2023) find when they increase the cumulative energy input per proton mass at mean energy, u_0. This suggests that heat injection might also be playing a role in the behaviour of the power spectrum in Figure 10. The value of u_0 in our simulations is likely to be higher than that in the reference simulation of Iršič et al. (2023) due to earlier reionization in our models.)

6 CONCLUSIONS

We have examined the effects of various assumptions and algorithms used in setting up the cosmological initial conditions for cosmological hydrodynamic simulations that are used to model the Lyman-α forest. Our goal was to understand if the predictions made by these models for the Lyman-α forest are robust against algorithmic choices and other assumptions. For this purpose, we designed simulations that were identical in all other respects, including in the manner of sampling the Gaussian initial density distribution. We considered five initial conditions codes, and studied four further variations of some of these codes. Our conclusions are as follows:

• At the high redshifts typically used to initialise cosmological simulations, when initial conditions codes employ a grid configuration for the initial unperturbed particle positions, their resultant perturbed density fields have a three-dimensional matter density power spectrum that agrees with the linear theory down to the smallest scales resolved in our simulations. This is not the case, however, when a glass-like configuration is used for the initial unperturbed particle positions. In these glass-based cases, the matter density power spectrum shows a significant enhancement at small scales relative to linear theory. This is also the case when a glass-like configuration is used only for gas particles.

• The non-linear dark matter and gas density power spectra, at redshifts $z = 5$–2, agree to within 10% in all our runs. For the dark matter density, the use of second-order Lagrangian perturbation theory causes the largest differences, relative to runs that use the Zel'dovich approximation. The use of adaptive softening can also lead to large differences in the dark matter density power spectrum. For gas, the largest difference is seen when a glass-like initial particle load is used. There are also large broadband differences when separate transfer functions are used for gas and dark matter. All of these differences, in dark matter and gas alike, decay with time, and are relatively insignificant at $z < 3$. The small-scale power enhancement can be offset by using adaptive softening.

• We report significant differences in the one-dimensional Lyman-α transmission power spectrum between our grid-based and glass-based runs. At $k \sim 0.1$ s/km, the flux power spectrum has differences of the order of 50% between our grid-based NGenIC run and our MP-GenIC run that uses a glass-like initial position for gas particles. These differences are scale-dependent, and therefore have a significant potential to bias any cosmological inferences made using similar models. These differences also appear to hold at high resolution. The source of the difference appears to be the choice of glass-like particle loads, which affects the small-scale gas velocities and densities. Of these two runs, only the MP-GenIC run correctly reproduces the expected linear theory behaviour in the difference of power between gas and dark matter densities.

It is not clear to us how important it is for models of the small-scale structure of the Lyman-α forest to reproduce the linear theory behaviour of the power difference between gas and dark matter densities. Nonetheless, the means of achieving this via the use of glass-like particle configurations appear to lead to large distortions in the Lyman-α flux power spectrum. These distortions can be considered a numerical artefact, for two reasons. First, these distortions appear similar to the distortions seen in fully glass-based runs, in which both

gas and dark matter particles are initialised on a glass, and where this behaviour is understood to be a numerical feature. Second, the matter density power spectrum in these runs fails to agree with the linear theory at even high redshifts. For all other things kept the same, this distortion causes an enhancement in the flux power spectrum at small scale. This can bias inferences in parameters such as the axion mass or the warm dark matter particle mass, and can potentially explain some of the very strict limits recently reported in the literature. In this context, it is interesting to note that Hahn et al. (2021) have also proposed a method to achieve the correct linear theory growth of the difference in the gas and dark matter density power spectra. In their approach the correct linear theory evolution is obtained by implementing the transfer functions as mass perturbation to each species, thus allowing the use of grids for both the dark matter and gas. This presents a potentially interesting alternative to consider in future efforts in the developing more robust models of the Lyman-α forest.

DATA AVAILABILITY

No new data were generated or analysed in support of this research.

REFERENCES

Angulo R. E., Springel V., White S. D. M., Jenkins A., Baugh C. M., Frenk C. S., 2012, MNRAS, 426, 2046
Armengaud E., Palanque-Delabrouille N., Yèche C., Marsh D. J. E., Baur J., 2017, MNRAS, 471, 4606

Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15

Baugh C. M., Gaztanaga E., Efstathiou G., 1995, MNRAS, 274, 1049

Baur J., Palanque-Delabrouille N., Yèche C., Boyarsky A., Ruchayskiy O., Armengaud É., Lesgourgues J., 2017, J. Cosmology Astropart. Phys., 2017, 013

Becker G. D., Bolton J. S., Haehnelt M. G., Sargent W. L. W., 2011, MNRAS, 410, 1096

Becker G. D., Hewett P. C., Worseck G., Prochaska J. X., 2013, MNRAS, 430, 2067

Bertschinger E., 1998, ARA&A, 36, 599

Bird S., Vogelsberger M., Sijacki D., Zaldarriaga M., Springel V., Hernquist L., 2013, MNRAS, 429, 3341

Bird S., Feng Y., Pedersen C., Font-Ribera A., 2020, J. Cosmology Astropart. Phys., 2020, 002

Boera E., Becker G. D., Bolton J. S., Nasir F., 2019, ApJ, 872, 101

Bolton J. S., Becker G. D., 2009, MNRAS, 398, L26

Bolton J. S., Puchwein E., Sijacki D., Haehnelt M. G., Kim T.-S., Meiksin A., Regan J. A., Viel M., 2017, MNRAS, 464, 897

Bosman S. E. I., et al., 2022, MNRAS, 514, 55

Chabanier S., et al., 2022, arXiv e-prints, p. arXiv:2207.05023

Crocce M., Pueblas S., Scoccimarro R., 2006, MNRAS, 373, 369

DESI Collaboration et al., 2016, arXiv e-prints, p. arXiv:1611.00036

Doughty C. C., Hennawi J. F., Davies F. B., Lukić Z., Oñorbe J., 2023, MNRAS,

Euclid Collaboration et al., 2019, A&A, 631, A85

Gaikwad P., et al., 2020, MNRAS, 494, 5091

Gaikwad P., Srianand R., Haehnelt M. G., Choudhury T. R., 2021, MNRAS, 506, 4389

Garaldi E., Gnedin N. Y., Madau P., 2019, ApJ, 876, 31

Garzilli A., Boyarsky A., Ruchayskiy O., 2017, Physics Letters B, 773, 258

Garzilli A., Magalich A., Theuns T., Frenk C. S., Weniger C., Ruchayskiy O., Boyarsky A., 2019, MNRAS, 489, 3456

Haardt F., Madau P., 2012, ApJ, 746, 125

Hahn O., Abel T., 2011, MNRAS, 415, 2101

Hahn O., Rampf C., Uhlemann C., 2021, MNRAS, 503, 426

Hiss H., Walther M., Oñorbe J., Hennawi J. F., 2019, ApJ, 876, 71

Hockney R. W., Eastwood J. W., 1981, Computer Simulation Using Particles

Iršič V., et al., 2017a, Phys. Rev. D, 96, 023522

Iršič V., Viel M., Haehnelt M. G., Bolton J. S., Becker G. D., 2017b, Phys. Rev. Lett., 119, 031302

Iršič V., Xiao H., McQuinn M., 2020, Phys. Rev. D, 101, 123518

Iršič V., et al., 2023, arXiv e-prints, p. arXiv:2309.04533

Karaçaylı N. G., et al., 2022, MNRAS, 509, 2842

Kim T. S., Bolton J. S., Viel M., Haehnelt M. G., Carswell R. F., 2007, MNRAS, 382, 1657

Kulkarni G., Hennawi J. F., Oñorbe J., Rorai A., Springel V., 2015, ApJ, 812, 30

LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201

Lee K.-G., et al., 2015, ApJ, 799, 196

Lesgourgues J., 2011, arXiv e-prints, p. arXiv:1104.2932

Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473

Lidz A., Faucher-Giguère C.-A., Dall'Aglio A., McQuinn M., Fechner C., Zaldarriaga M., Hernquist L., Dutta S., 2010, ApJ, 718, 199

Lukić Z., Stark C. W., Nugent P., White M., Meiksin A. A., Almgren A., 2015, MNRAS, 446, 3697

Maiolino R., et al., 2013, arXiv e-prints, p. arXiv:1310.3163

McDonald P., Miralda-Escudé J., Rauch M., Sargent W. L. W., Barlow T. A., Cen R., Ostriker J. P., 2000, ApJ, 543, 1

McQuinn M., 2016, ARA&A, 54, 313

Meiksin A. A., 2009, Reviews of Modern Physics, 81, 1405

Molaro M., et al., 2022, MNRAS, 509, 6119

O'Leary R. M., McQuinn M., 2012, ApJ, 760, 4

Palanque-Delabrouille N., Yèche C., Schöneberg N., Lesgourgues J., Walther M., Chabanier S., Armengaud E., 2020, J. Cosmology Astropart. Phys., 2020, 038

Peacock J. A., Heavens A. F., 1985, MNRAS, 217, 805

Pieri M. M., et al., 2016, in Reylé C., Richard J., Cambrésy L., Deleuil M., Pécontal E., Tresse L., Vauglin I., eds, SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. pp 259–266 (arXiv:1611.09388)

Planck Collaboration XVI 2014, A&A, 571, A16

Puchwein E., et al., 2023, MNRAS, 519, 6162

Ricotti M., Gnedin N. Y., Shull J. M., 2000, ApJ, 534, 41

Rogers K. K., Peiris H. V., 2021, Phys. Rev. Lett., 126, 071302

Rollinde E., Theuns T., Schaye J., Pâris I., Petitjean P., 2013, MNRAS, 428, 540

Rorai A., et al., 2017a, Science, 356, 418

Rorai A., et al., 2017b, MNRAS, 466, 2690

Rorai A., Carswell R. F., Haehnelt M. G., Becker G. D., Bolton J. S., Murphy M. T., 2018, MNRAS, 474, 2871

Salmon J., 1996, ApJ, 460, 59

Schaye J., Theuns T., Rauch M., Efstathiou G., Sargent W. L. W., 2000, MNRAS, 318, 817

Scoccimarro R., 1998, MNRAS, 299, 1097

Scoccimarro R., Hui L., Manera M., Chan K. C., 2012, Phys. Rev. D, 85, 083002

Skidmore W., TMT International Science Development Teams Science Advisory Committee T., 2015, Research in Astronomy and Astrophysics, 15, 1945

Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79

Springel V., et al., 2005, Nature, 435, 629

Theuns T., Schaye J., Zaroubi S., Kim T.-S., Tzanavaris P., Carswell B., 2002, ApJ, 567, L103

Viel M., Lesgourgues J., Haehnelt M. G., Matarrese S., Riotto A., 2005, Phys. Rev. D, 71, 063534

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., Rauch M., Sargent W. L. W., 2008, Phys. Rev. Lett., 100, 041304

Viel M., Becker G. D., Bolton J. S., Haehnelt M. G., 2013, Phys. Rev. D, 88, 043502

Villasenor B., Robertson B., Madau P., Schneider E., 2021, ApJ, 912, 138

Villasenor B., Robertson B., Madau P., Schneider E., 2022, ApJ, 933, 59

Villasenor B., Robertson B., Madau P., Schneider E., 2023, Phys. Rev. D, 108, 023502

Walther M., Hennawi J. F., Hiss H., Oñorbe J., Lee K.-G., Rorai A., O'Meara J., 2018, ApJ, 852, 22

Walther M., Oñorbe J., Hennawi J. F., Lukić Z., 2019, ApJ, 872, 13

Walther M., Armengaud E., Ravoux C., Palanque-Delabrouille N., Yèche C., Lukić Z., 2021, J. Cosmology Astropart. Phys., 2021, 059

White S. D. M., 1996, in Schaeffer R., Silk J., Spiro M., Zinn-Justin J., eds, Cosmology and Large Scale Structure. p. 349

Wu X., McQuinn M., Kannan R., D'Aloisio A., Bird S., Marinacci F., Davé R., Hernquist L., 2019, MNRAS, 490, 3177

Zaldarriaga M., Hui L., Tegmark M., 2001, ApJ, 557, 519

Zel'dovich Y. B., 1970, A&A, 5, 84

APPENDIX A: PARTICLE OFFSETS IN INITIAL CONDITIONS

As discussed above, in Section 2.2.1, we introduce an offset in gas particle positions relative to dark matter particle positions in our NGenIC run. This is done to avoid
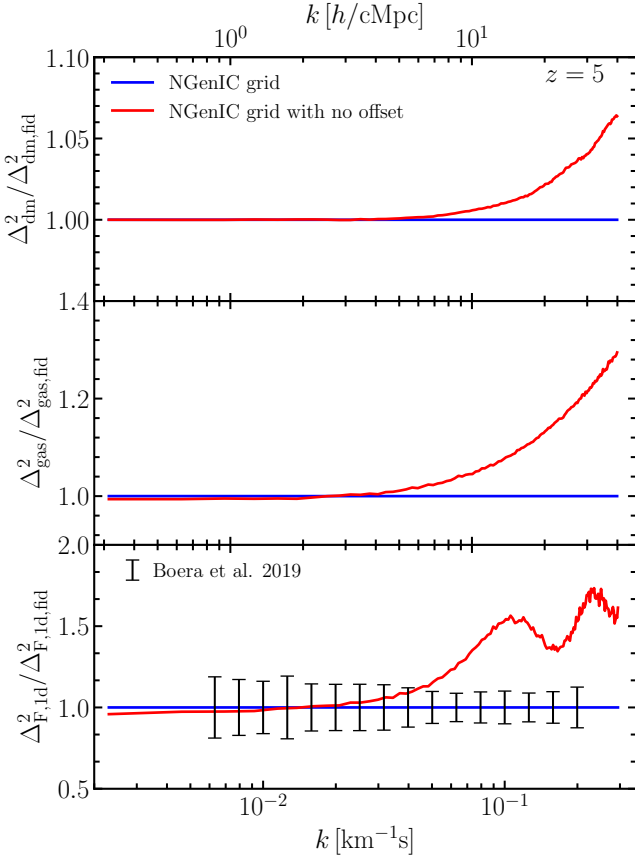
Figure B1. The one-dimensional Lyman-α transmission power spectrum at $z = 5$ in a simulation that does not account for the radiation density term (blue curve) and one that does account for it (red curve). Both simulations use NGenIC with an initial power spectrum derived using CLASS and normalised to give $\sigma_8 = 0.829$ at $z = 0$.

APPENDIX B: EFFECT OF RADIATION

We use CLASS to obtain the initial power spectrum at $z = 99$. This input power spectrum is normalised to obtain our desired value of $\sigma_8 = 0.829$ at $z = 0$. This normalisation is done using a growth factor that does not include the radiation term, in order to be consistent with our use of the P-Gadget3 code, which also does not include the radiation term while evolving the density perturbations. Figure B1 shows the effect that this has on the Lyman-α flux power spectrum relative to a P-Gadget3 run that uses the radiation term for the background cosmological evolution for an identical input power spectrum. The figure shows the power spectrum from two 20–256 simulations using P-Gadget3, one of which does not use the radiation density term (blue curve) and one that does use it (red curve). Both simulations use our default grid-based NGenIC set-up to compute the initial conditions, so that the input power spectrum is normalised in identical fashion in both cases. We see that an inconsistency in incorporating the radiation term leads to a small bias of less than 10% in the power at $k \sim 0.1$ s/km. There is no significant effect at larger scales.

APPENDIX C: STARTING REDSHIFT

Figure C1 shows the effects of changing the starting redshift of our simulations from $z = 99$ to $z = 199$ and $z = 49$. This has a very small impact on the result. The small differences that do exist seem to indicate that the difference between the power spectra in the MP-GenIC and NGenIC cases are smaller for a smaller starting redshift. This observation, combined with the similarity in the shape of the distortion seen in the bottom panel of Figure A1 with that in Figure 7 and other related figures, suggests that the distortions may arise at least partly due to particle coupling. When the simulation is started at a later redshift there is relatively less time for the accumulation of particle coupling error, which explains the behaviour seen in Figure C1.

Figure A1. Effect of using a position offset in the initial gas and dark matter particle positions on the three-dimensional dark matter density power spectrum (top panel), the three-dimensional gas density power spectrum (middle panel), and the one-dimensional Lyman-α transmission power spectrum (bottom panel). We plot the ratio of the power spectrum from a simulation that does not offset the particle positions, thus placing gas particles at the same positions as the dark matter particles while starting the simulation. In the blue curve, the default offset, explained in Section 2.2.1, is used. Both simulations are 20–256 and use NGenIC.

spurious small-scale power due to gravitational coupling between the two species of particles. Figure A1 shows the effect that not offsetting the particles has on the dark matter density power spectrum, the gas density power spectrum, and the one-dimensional Lyman-α transmission power spectrum. Only redshift $z = 5$ is shown for illustration. As noted by O'Leary & McQuinn (2012), gravitational coupling between gas and dark matter particles can lead to a significant excess power at small scales. It is interesting to note that particle offsets in a grid-based NGenIC run produce changes in the transmission power spectrum that are similar to the changes introduced when one goes from the grid-based NGenIC run with offset to the MP-GenIC run (Figure 6). This suggests that the differences between MP-GenIC and NGenIC could be caused by particle coupling and two-body interactions.
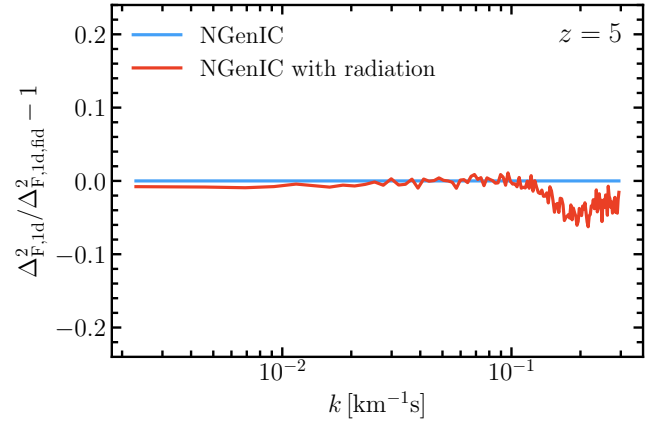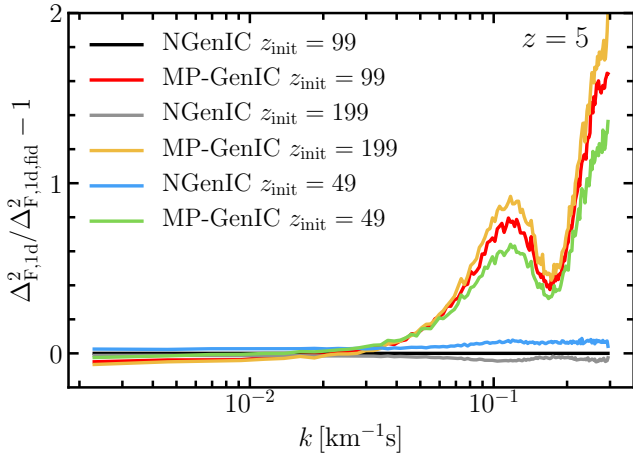
Figure C1. Dependence of our results on the starting redshift of the simulations. The black and red curves show results from our default NGenIC and MP-GenIC simulations, respectively, both of which start at $z_{\mathrm{init}} = 99$. The grey and orange curves show results from the simulation when they are initialised at $z_{\mathrm{init}} = 199$ instead. The blue and green curves show the same for $z_{\mathrm{init}} = 49$. Changing the initialisation redshift does not have a large effect on the flux power spectrum.

APPENDIX D: SOFTENING LENGTH CONVERGENCE TEST

Figures D1 and D2 show how the result in Figure 10 depends on the choice of the values of the `PMGRID` and l_{soft} parameters. The `PMGRID` parameter controls the resolution of the mesh used in the PM force calculation. The softening length l_{soft} controls the short-range interaction between particles. The bottom panels of Figures D1 show that the results of our 20–1024 simulation are not very sensitive to the value of `PMGRID`. The same is true for the value of l_{soft} for the NGenIC run. However, the MP-GenIC run shows significant changes of the order of 10% in the flux power spectrum. Figure D2 shows that the difference in the power spectra in the two simulations worsens with decreasing softening length. The reduction in the difference between NGenIC and MP-GenIC with increasing softening length suggests that the differences could be caused by particle coupling and two-body interactions.

APPENDIX E: ONE-DIMENSIONAL Lyman-α TRANSMISSION PDF

The Lyman-α flux probability distribution function (PDF) is shown in Figure E1, with the differences relative to the grid-based NGenIC run plotted at the bottom of each panel. We also show measurements of the PDF by McDonald et al. (2000), Kim et al. (2007), and Rollinde et al. (2013). To compute the PDF, we took 50 equal-sized bins between 0 and 1. At $z = 4$ and 5, the flux PDF peaks at $F = 0$ but at lower redshifts the mode moves to $F = 1$ as the forest becomes increasingly transparent. Like all the other statistics, the difference between the simulations for Lyman-α transmission PDF decreases with time. The differences are highest at $z = 5$. At these redshifts the PDF from the MP-GenIC run shows almost 100% difference near $F = 1$ from that in

the grid-based NGenIC run. The glass-based NGenIC run is discrepant too, but with smaller differences, of about 20% at $F = 1$. Interestingly, using adaptive softening seems to enhance these differences. There is good agreement between the rest of the simulation, except perhaps in the case of 2LP-TIC, which also shows about 20% differences, most likely as a result of the higher accuracy of the density perturbation computation.
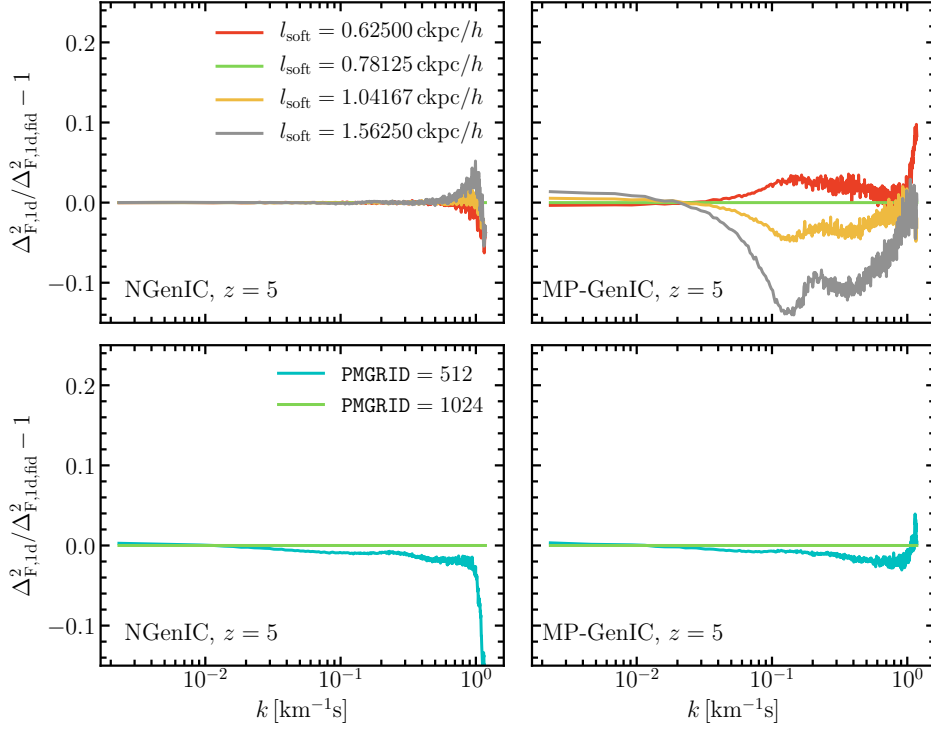
Figure D1. Gravitational softening length and `PMGRID` convergence tests for NGenIC and MP-GenIC at redshift $z = 5$. All four panels show the change in the one-dimensional Lyman-α flux power spectrum with respect to one of the default run. This is shown by the green curves in all the panels, and represents our default choice of $l_{\rm soft} = 0.78125\,{\rm ckpc}/h$ and `PMGRID` = 1024. The left panels show the results for the NGenIC and the right panels are for the MP-GenIC initial condition generators. The red, orange and grey curves show the results for softening lengths of 0.80, 1.33 and 2 times the default value keeping the `PMGRID` value to the default. From the upper two panels, we can see that for NGenIC the power spectrum is converged with respect to the softening length, but that from the MP-GenIC had not. MP-GenIC produces more power with reduced softening. In the lower panels, we use `PMGRID` = 512 for the cyan-colored curves. The Lyman-α flux power spectrum does not change significantly with `PMGRID` in either of the codes.
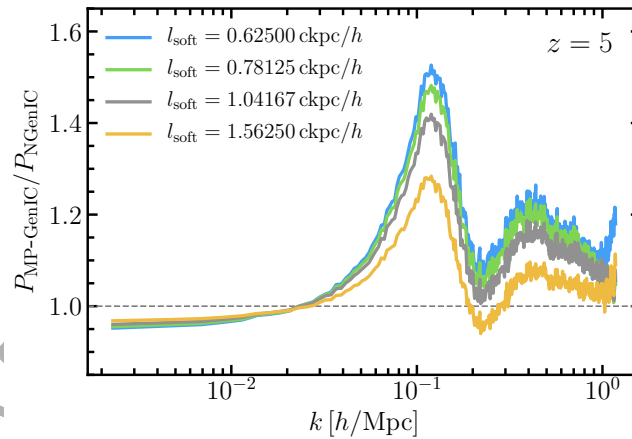

Figure D2. Dependence of the Lyman-α flux power ratio between MP-GenIC and NGenIC on the gravitational softening length. We keep `PMGRID` fixed to the fiducial value of 1024 in all the curves here. The default softening length is shown by the green curve. The blue, grey and orange curves are for softening lengths of 0.8, 1.33 and 2 times the default length. Reducing the softening length increases the differences between the two codes.
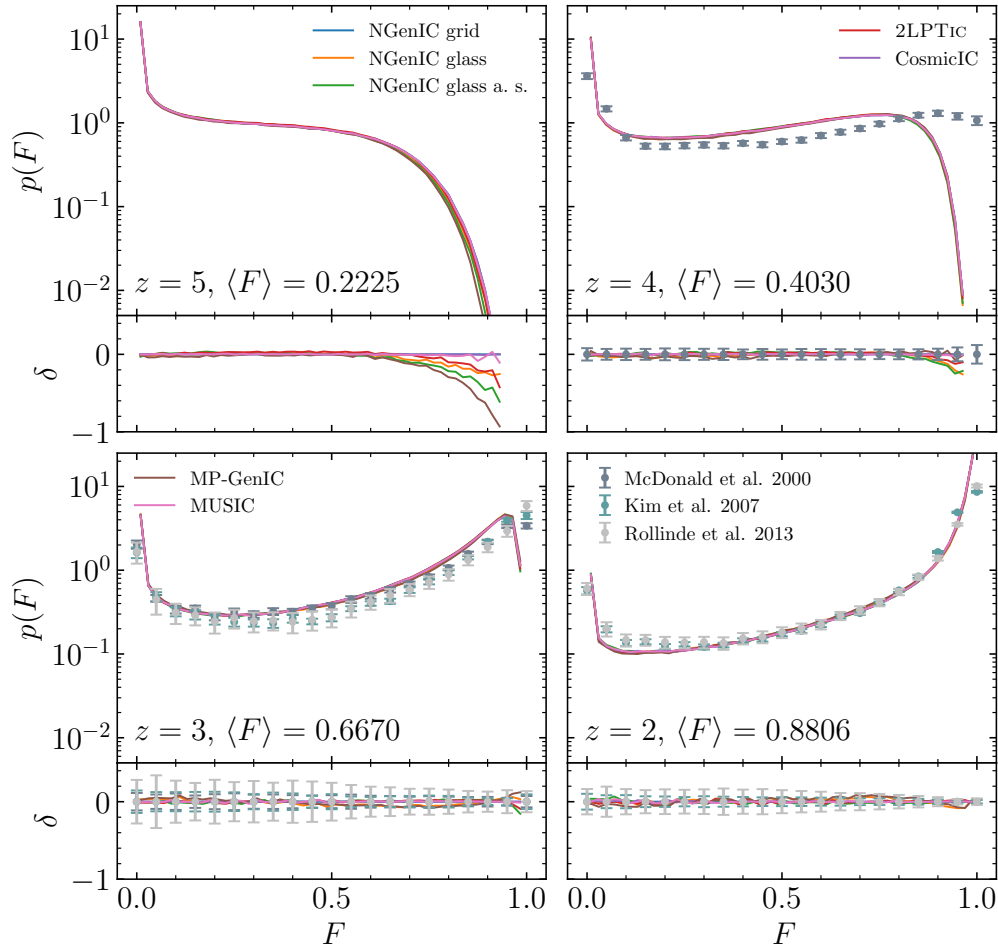
Figure E1. The PDF of Lyman-α transmission at $z = 2, 3, 4$, and 5. Each panel shows results from our five primary simulations. Also shown are results from a glass-based NGenIC run, and a glass-based NGenIC run that uses adaptive softening (indicated by 'a. s.' in the legend). We also show measurements by McDonald et al. (2000), Kim et al. (2007), and Rollinde et al. (2013). These data are only shown for context and as a measure of the current uncertainties. We make no attempt at fitting the data with our models, which explains the mismatch between the simulations and the observations at some redshifts and transmission values. The error panels show differences relative to the result from the grid-based NGenIC simulation. All spectra are normalised to the observed mean transmission, the values of which are given in the respective panels.