How to Zoom: Bias, Contamination, and Lagrange Volumes in Multimass Cosmological Simulations

Jose Oñorbe,1,2* Shea Garrison-Kimmel,1 Ariyeh H. Maller,3,4 James S. Bullock,1,2 Miguel Rocha,1 and Oliver Hahn5

1 Center for Cosmology, Department of Physics and Astronomy, The University of California at Irvine, Irvine, CA, 92697, USA
2 Center for Galaxy Evolution, Department of Physics and Astronomy, The University of California at Irvine, Irvine, CA, 92697, USA
3 Department of Physics, New York City College of Technology, CUNY, Brooklyn, NY 11201, USA
4 Department of Astrophysics, American Museum of Natural History, New York, NY 10024, USA
5 ETH Zurich Institute for Astronomy, CH-8093 Zurich, Switzerland

ABSTRACT

We perform a suite of multimass cosmological zoom simulations of individual dark matter halos and explore how to best select Lagrangian regions for resimulation without contaminating the halo of interest with low-resolution particles. Such contamination can lead to significant errors in the gas distribution of hydrodynamical simulations, as we show. For a fixed Lagrange volume, we find that the chance of contamination increases systematically with the level of zoom. In order to avoid contamination, the Lagrangian volume selected for resimulation must increase monotonically with the resolution difference between parent box and the zoom region. We provide a simple formula for selecting Lagrangian regions (in units of the halo virial volume) as a function of the level of zoom required. We also explore the degree to which a halo’s Lagrangian volume correlates with other halo properties (concentration, spin, formation time, shape, etc.) and find no significant correlation. There is a mild correlation between Lagrange volume and environment, such that halos living in the most clustered regions have larger Lagrangian volumes. Nevertheless, selecting halos to be isolated is not the best way to ensure inexpensive zoom simulations. We explain how one can safely choose halos with the smallest Lagrangian volumes, which are the least expensive to resimulate, without biasing one’s sample.

Key words: cosmology: theory — galaxies: formation — galaxies: halos — methods: N-body simulations

1 INTRODUCTION

Over the past three decades, numerical simulations of cosmic structure formation have become a standard tool for studying a wide range of problems in cosmology. The predictive power of such simulations increases with their complexity, but even with the increasing size and computational power of the machines used to run them, advances in computational algorithms have played a key role in pushing the limits of this method. For example, the multimass technique (Porter 1985; Katz & White 1993; Navarro & White 1994) has converged as a common approach to study the role of physical processes important to galaxy formation in a cosmological context.

The basic method behind this “zoom” technique involves selecting a specific interesting region from a low resolution simulation and re-running it with higher resolution particles in that region. In this way, computational resources are focused mainly on the area of interest while the long range forces of gravity are still captured in their appropriate cosmological context (Navarro & White 1994; Frenk et al. 1996; Tormen et al. 1997; Thacker & Couchman 2000; Bertschinger 2001; Klypin et al. 2001; Power et al. 2003; Springel et al. 2008; Jenkins 2010; Hahn & Abel 2011). Currently, this technique is widely used to study the formation and evolution of all types of objects in a cosmological context, from dwarf galaxy halos to large clusters.

The region of interest is usually the virialized region of a dark matter halo, though, depending on the purpose of the simulation, it can be larger (e.g., Hahn et al. 2010; Oñorbe et al. 2011; Garrison-Kimmel et al. 2013). Once the region of interest is selected, particles in that volume are traced back to their initial positions, thus defining a Lagrange volume (so called because its key feature is that the particles trace the flow of matter through cosmic time). One then creates initial conditions with higher spatial and mass resolution in the Lagrange volume, while elsewhere retaining the lower resolution of the original box, such that long range forces from surrounding matter are properly resolved at minimal CPU and memory cost. In addition, several boundary or buffer regions are typi-
cally initialized at intermediate resolution around the Lagrange volume. This method allows one to use a periodic box large enough to obtain cosmological convergence, including low frequency modes (Power & Knebe 2006), while simultaneously reaching resolutions high enough to study the small scale processes relevant to galaxy formation.

Generating initial conditions for such a simulation requires the following steps:

i: Determine the box size and resolution required in both the full- and multimass simulations to study the given problem. 

ii: Select a cosmological model ($\Omega_m, \Omega_\Lambda, \sigma_8$, etc.), calculate the appropriate initial redshift, and create a matter power spectrum (or transfer function) for those parameters at that redshift to generate the initial density fluctuations.

iii: Identify the region of interest in which higher resolution is desired. Usually this region is selected from a lower resolution full-box simulation by selecting a certain volume around a galactic halo. Particles inside this volume are identified and traced back to the initial conditions, and the Lagrange volume is then defined by the initial positions of those particles.

iv: Discretize the mass distribution in the simulation, according to the desired resolution, with the appropriate density fluctuations generated using the initial redshift matter power spectrum.

Motivated primarily by the concern that the original approaches might be too simplistic and not accurate enough for the precision needed in full-box and multimass cosmological simulations, the preceding steps have been extensively refined over the last several years. Several authors have contributed to this effort to a varying degree of detail; see, e.g. [Navarro & White (1994; Frenk et al. 1996; Tormen et al. 1997; Thacker & Couchman 2000; Bertschinger 2001; Klypin et al. 2001; Power et al. 2003; Springel et al. 2008; Jenkins 2010; Hahn & Abel 2011].

In particular, theorists have put a great deal of effort into improving the final step listed above which applies to both full-box and multimass simulations. These efforts fall into two categories: first, simulators have worked to properly generate the perturbations of the distribution from the overdensity field using second order perturbation theory (Sirko 2005; Joyce et al. 2009; Jenkins 2010) and second, they have improved the numerical techniques used to set the discrete distribution of mass, in particular by covolvoling the transfer function with (typically Gaussian) white noise to generate the correct overdensity field (Bertschinger 2001; Hahn & Abel 2011).

However, many of the choices that one must make in order to create multimass initial conditions have not been discussed before in the literature. For example: How large does the Lagrange volume need to be in order to avoid contamination in the high-resolution region by low-resolution particles? Can one choose halos with intrinsically small Lagrange volumes in order to save cost without biasing results? At what redshift must one start the resimulation? How many buffer regions are necessary and how large should they be? Most groups who perform multimass simulations have arrived at their methodologies by trial and error and/or the wisdom of experience. In this paper, we improve upon this situation by testing the effects of these choices and quantifying their relative importance. The paper is organized as follows. In Section 2 we discuss the simulations we have run and the tools we have used to analyze them. In Section 3 we discuss methods for identifying the Lagrangian region of interest. In Section 4 we discuss how to avoid low-resolution particle contamination in the high-resolution region. Section 5 shows that the Lagrange volume of a halo does not correlate with other halo properties except that objects in high-density environments tend to have higher Lagrange volume. The implication is that halos with small Lagrange volume regions can be chosen for resimulation without biasing the outcome. Section 6 discusses how the initial redshift of the zoom simulation and buffer regions around the Lagrange volume should be chosen. We close with conclusions in Section 7.

| Name               | $L_{box}$ ($h^{-1}$Mpc) | $N_p$ | $m_p$ ($M_\odot$) | N |
|--------------------|------------------------|------|-------------------|---|
| L5n256             | 5                      | 256^4 | 7.7 x 10^7         | 2 |
| L5n512             | 5                      | 512^4 | 9.6 x 10^4         | 1 |
| L25n128            | 25                     | 128^3 | 7.7 x 10^8         | 1 |
| L25n256            | 25                     | 256^3 | 9.6 x 10^7         | 1 |
| L25n512            | 25                     | 512^3 | 1.2 x 10^7         | 3 |
| L50n128            | 50                     | 128^3 | 6.1 x 10^9         | 1 |
| L50n256            | 50                     | 256^3 | 7.7 x 10^8         | 8 |
| L50n512            | 50                     | 512^3 | 9.6 x 10^7         | 3 |
| L150n256           | 150                    | 256^3 | 2.1 x 10^10        | 1 |
| L650n128           | 650                    | 128^3 | 1.4 x 10^13        | 1 |
| L650n256           | 650                    | 256^3 | 1.8 x 10^12        | 1 |
| L650n512           | 650                    | 512^3 | 2.2 x 10^11        | 2 |
| L900n512           | 900                    | 512^3 | 5.6 x 10^11        | 1 |

\[ ^{a} \text{Also ran at varying } z_{\text{ini}} \]
\[ ^{b} \text{Also ran with ENSO} \]
\[ ^{c} \text{Adiabatic version also run using GADGET-2} \]

2 SIMULATIONS

We have run a series of cosmological N-body simulations following the formation and evolution of structure in the ΛCDM model. We use a sequence of boxes of side length $L_{box}$ that vary between 5 and 900 $h^{-1}$Mpc. In each case we assume a flat cosmology with a cosmological constant term based on the latest results of WMAP (Komatsu et al. 2011), with cosmological parameters \( \Omega_\Lambda = 0.734 \), \( \Omega_m = 0.266 \), \( \Omega_b = 0.0449 \), \( h = 0.71 \), normalization of an initial power-law index \( n = 0.963 \) and \( \sigma_8 = 0.801 \).

In order to generate the initial conditions (ICs), we have used the MUSIC code (Hahn & Abel 2011). The method uses an adaptive convolution of Gaussian white noise with a real-space transfer function kernel together with an adaptive multi-grid Poisson solver to generate displacements and velocities following second-order Lagrangian perturbation theory. For more specific details on the MUSIC code, we kindly refer the reader to (Hahn & Abel 2011). The transfer functions used to generate the ICs for this cosmology were obtained using CAMB (Lewis et al. 2000; Howlett et al. 2012). The initial redshift, $z_{\text{ini}}$, used for all our simulations is discussed in Section 6.

To run the simulations in this work we used the publicly available codes GADGET-2, a Lagrangian TreeSPH code (Springel et al. 2008; Jenkins 2010; Hahn & Abel 2011).
Table 2. Zoom-in simulations. First column: full simulation from which the resimulated halo was chosen and the Lagrange volume was calculated. Second column: Virial mass of the simulated halo at \( z = 0 \). Third column: the number of different halos that we simulated at that mass, if more than one. Fourth column: Number of zoom-in simulations run with GADGET-2 using the cuboid Lagrange volume approach (including sixteen adiabatic simulations). Fifth column: Number of zoom-in simulations run with ENZO using the cuboid Lagrange volume approach. Sixth column: Total number of zoom-in simulations. These simulations include runs with varying padding, resolution, \( R_{18h} \) definition, etc. See text for more details.

| Full Box | \( M_V \) \((M_{⊙})\) | Halos | \( N_G \) | \( N_E \) | Total |
|----------|-----------------|------|------|------|------|
| L5n256   | \( 5 \times 10^9 \) | 1    | 3    | 0    | 3    |
| L5n512a  | \( 1 \times 10^{10} \) | 1    | 2    | 0    | 2    |
| L5n512a  | \( 5 \times 10^9 \) | 1    | 10   | 0    | 10   |
| L5n512a  | \([0.8-1.2] \times 10^{10}\) | 2    | 7    | 0    | 7    |
| L5n512a  | \( 4 \times 10^{10} \) | 1    | 4    | 0    | 4    |
| L5n512   | \([3-6] \times 10^{11}\) | 3    | 19   | 25   | 44   |
| L5n512   | \([0.8-1.2] \times 10^{12}\) | 23   | 88   | 2    | 90   |
| L5n512   | \( 5 \times 10^{12} \) | 1    | 11   | 2    | 13   |
| L65n512  | \([0.8-1.2] \times 10^{15}\) | 4    | 8    | 0    | 8    |
|          |                  | 152  | 29   | 181  |      |

- \( a_{50} \): Time of formation
- \( N_{subh} \): Number of subhalos
- \( N_{mergers} \): Number of major mergers
- \( z_{in} \): Redshift of last major merger
- \( N_{neigh} \): Number of nearby halos
- \( D_{1,X} \): Distance to nearest massive halo
- \( M_{gas}/M_V \): Gas fraction
- \( \lambda_{gas} \): Gas spin parameter
- \( V \): Cuboid Lagrange volume
- \( V_{LH} \): Convex hull Lagrange volume
- \( V_{MC} \): Minimum cuboid Lagrange volume
- \( V_{ME} \): Minimum ellipsoid Lagrange volume
- \( R_{tb} \): Traceback radius of the Lagrange volume
- \( (w/l)\nu_{MC} \): Lagrange volume cubic measure
- \( \Delta_{res} \): Difference in resolution between the zoom and the parent volume
- \( z_{ini} \): Initial redshift

The gravitational softening used for these fiducial full-box simulations was set to \( \epsilon = 0.02 \times L_{box} / \sqrt{N_p} \). In addition, four combinations of box size and particle number (\( L_{5n256}, L_{50n256}, L_{50n512}, \) and \( L_{65n512} \)) were run with a different initial redshift \( z_{ini} \). Two cases (\( L_{5n256} \) and \( L_{50n512} \)) were run with different force resolutions. Finally, an adiabatic version of two boxes (\( L_{50n256} \) and \( L_{50n512} \)) were also run.

From these full-box runs, we have selected regions of interest and run 181 zoom-in simulations, summarized in Table 2. We used the Amiga Halo Finder (AHF) \(^{(2005)}\), and ENZO, a hybrid adaptive mesh refinement (AMR), grid-based code \(^{(2004)}\). We duplicate our results with these two codes not only to confirm that findings concerning the Lagrange volume hold for both approaches, but also to study the possible differences between Lagrange and Eulerian approaches to creating initial conditions for multimass simulations.

We have run a total of 26 full-box simulations. These consist of thirteen different combinations of \( L_{box} \) and particle number \( N_p \) (or mass, \( m_p \)), summarized with an associated naming convention in Table 3. The gravitational softening used for these fiducial full-box simulations was set to \( \epsilon = 0.02 \times L_{box} / \sqrt{N_p} \). In addition, four combinations of box size and particle number (\( L_{5n256}, L_{50n256}, L_{50n512}, \) and \( L_{65n512} \)) were run with a different initial redshift \( z_{ini} \), while two cases (\( L_{25n512} \) and \( L_{50n512} \)) were run with different force resolutions. Finally, an adiabatic version of two boxes (\( L_{50n256} \) and \( L_{50n512} \)) were also run.

From these full-box runs, we have selected regions of interest and run 181 zoom-in simulations, as summarized in Table 2. The regions to zoom-in on were selected to contain a range of halo masses. We followed Power et al. \(^{(2003)}\) in setting the force resolutions for the multimass simulations: \( \epsilon = 4 \times R_{vir} / \sqrt{\chi} \), where \( \chi \) is the number of particles expected in the virial radius (based on the full-box-derived virial mass and resimulated particle mass). These simulations were run with varying mass resolutions and definitions of the Lagrange volume, as discussed in the sections to follow. The mass resolution used in these runs range from an effective resolution of \( 512^3 \) to \( 4096^3 \) particles in the high resolution volume. We also ran sixteen adiabatic multimass simulations with GADGET-2 of a Milky Way halo in which an identical softening was used for the gas particles.

2.1 Halo Identification and Analysis

We used the Amiga Halo Finder (AHF) \(^{(2009)}\) to identify halos and calculate most of the halo properties, and developed our own pipeline to calculate any property not determined by AHF. Unless specifically stated, this pipeline is based on parameters given by AHF (center, virial radius, etc). The virial mass (\( M_V \)) is calculated using the overdensity (\( \Delta_{vir} \)) formula from Bryan & Norman \(^{(1998)}\) for our cosmology at each specific redshift. Our conclusions do not change when using different overdensity definitions, e.g. \( \Delta_{200} = 200 \rho_{crit} \). Note that we only consider distinct halos throughout this study as the Lagrange volume of a subhalo is set by its host.

A list of the main halo properties used in this paper and their symbols are listed in Table 3. Throughout, we use the Bullock et al. \(^{(2001)}\) definition for \( \lambda \) and the standard definition for triaxiality, \( T \equiv (a^2 - b^2)/ (a^2 - b^2) \) where \( a, b, c \) are the square root of the eigenvalues of the moment of inertia tensor \(^{(2006)}\). We also calculated both the number of major mergers since \( z = 1 \) (\( N_{mergers} \)) and over all time (\( N_{mergers} \)), where we define a major merger by a mass ratio greater than 0.4 (although other values were also tested). We quantify formation time by the scale factor (\( a_{50} \)) or redshift (\( z_{50} \)) at which the virial mass of the halo reaches half of the \( z = 0 \) virial mass. We also calculate two environmental parameters: \( N_{neigh}^{(3)} \) following Haas et al. \(^{(2012)}\), which is the number of halos outside the virial radius of a halo but within a sphere of radius 3 Mpc from the halo center with a mass ratio greater than 0.3 (other values for the mass ratio and the radius were also tested); and \( D_{1,0.1} \), which is the distance from the center of a halo to the nearest neighbor with a mass ratio greater than 0.1 (other values were also tested), divided by \( R_V \) of the neighboring halo. As Jesseon-Daniel et al. \(^{(2011)}\) and Haas et al. \(^{(2012)}\) have shown, the first parameter correlates very strongly with virial mass,
Figure 1. Lagrange volume definitions. In the left panel, the red points indicate the $z = z_{\text{ini}}$ positions of all the particles (1510) within $R_v$ at $z = 0$ of a halo with $M_v = 1.46 \times 10^{11} M_\odot$ (from L50n512). The black cube stands for the cuboid Lagrange volume, $V$, i.e., the cuboid that is aligned with the simulation axis and that encloses all the particles. The green cube shows the minimum cuboid that encloses the selected particles ($V_{MC}$). The transparent grey ellipsoid shows the minimum ellipsoid that encloses the selected particles ($V_{ME}$). Note the majority of the particles at $z = z_{\text{ini}}$ are nearly planar; we find that this is a generic feature of the Lagrange volumes we explore. The right panel again shows $V$ (black cube), along with the convex hull Lagrange volume, $V_{CH}$, for the same halo. As expected, the convex hull volume is much closer to the actual particle distribution.

whereas the second is insensitive to mass. We find similar results as Jeeson-Daniel et al. (2011) and Skibba & Macciò (2011) regarding the correlations between these parameters, and therefore do not investigate these correlations further.

3 IDENTIFYING THE LAGRANGE VOLUME

Identifying the Lagrange volume that one will resimulate at high resolution is an important step in the multimass technique. For this task, one first selects a set of particles at the redshift of interest (e.g. particles within $R_v$, or some multiple of $R_v$, at $z = 0$), then determines a volume that contains all of those particles at $z = z_{\text{ini}}$. However, neither step is precisely defined, and techniques used in the literature vary; here, we explore those different definitions. We first investigate the relationship between the different volumes that may be identified at $z = z_{\text{ini}}$ for a fixed set of particles, then we discuss the effects of different methods for selecting the particle set at the redshift of interest (e.g. $z = 0$). Throughout this paper, we quote the Lagrange volume in comoving units of $h^{-1}\text{Mpc}^3$.

We consider four Lagrange volume definitions for particles: the cuboid volume $V$, the minimum rotated cuboid volume $V_{MC}$, the minimum ellipsoid $V_{ME}$, and the convex hull volume $V_{CH}$. The four choices are illustrated in Figure 1. The cuboid volume (black boxes in Figure 1) is the minimum cuboid in $(x,y,z)$ coordinates aligned with the axes of the simulation that contains the set of particles at $z = z_{\text{ini}}$. This is the most straightforward volume to define, set by the minimum and maximum values of $x$, $y$, and $z$ for the chosen particles. This volume is the largest of the cases we consider, but is more practical for many codes and/or initial conditions generators. The convex hull (Barber et al. 1996) is the minimal polyhedron that encloses the selected particles at $z = z_{\text{ini}}$. This choice provides the minimum volume and is illustrated relative to the cuboid in the right panel of Figure 1. The minimum rotated cuboid, an intermediate choice, is illustrated by the green cube in Figure 1. The transparent grey ellipsoid in the same Figure that encloses all the selected particles is the minimum rotated ellipsoid (Khachiyan 1979). It is important to note that the convex hull is the best physical description of the Lagrange volume and removes unphysical effects such as the random alignment between the cuboid and the simulation axis; however, while this alignment has no physical significance, in practice it is not trivial to rotate the simulation and thus this issue cannot be ignored. As a reference, Table 3 summarizes the symbols of the main Lagrange volume parameters used in this work.

We also analyze the effect of varying the particle set that defines the Lagrange volume. One natural choice is to pick a Lagrange volume set by the particles within one virial radius of a halo. However, such a definition may lead to contamination within the virial radius caused by wandering low resolution particles that can end up within the virial volume (see Section 4.2 for a discussion of the effects of contamination); thus, one typically selects particles from a larger radius than the region of interest in the final zoom. How conservative one must be in selecting the volume for resimulation is one of the points we return to later. We explore a few choices of “traceback” radius, $R_{tb}$, for defining the Lagrangian region of interest. First we consider $R_{tb}$ defined as a multiple of the halo’s $z = 0$ virial radius. Another option, which we explore later in the paper, is to use merger trees to track all particles that have ever been part of the halo of interest (Oser et al. 2010; Feldmann et al. 2010). We apply this approach in two ways. First, in the “main branch” (mb) method, we select all particles that are ever within $(1, 3) \times R_v(z)$ of the main halo; and second, in the “full merger tree” (mt) method, we select all particles that are ever
within $(1, 3) \times R_v(z)$ of any halo that ends in the $z = 0$ object. For now, we will focus on the simplest definition set by the $z = 0$ particle positions. We show below that the more computationally expensive approaches that rely on merger trees do not help the contamination issue appreciably.

Figure 2 provides an illustration of how Lagrange volume scales with halo mass. As expected, there is a strong and intuitive correlation such that more massive halos have larger Lagrange volumes. To capture any unexpected trends, we study the normalized correlation such that more massive halos have larger Lagrange volumes when normalized by $R_v$. Results from L5n512 are affected by the small size of the box and therefore were removed from the fits and are shown using empty squares. The lines become dashed where the fit is extrapolated.

Figure 3. Relative sizes of the Lagrange volume definitions as a function of halo mass. The volumes are defined by the particles within a traceback radius $R_{tb} = 1 R_v$ of the halo center at $z = 0$ (though the results are largely independent of $R_{tb}$). Symbols stand for the median value of each mass bin and error bars show the 25% and 75% percentiles. We plot the ratios of the cuboid ($V^1$ black), rotated cuboid ($V^1_{MC}$ green) and ellipsoid ($V^1_{ME}$ blue) Lagrange volumes to the convex hull definition calculated in the $N_p = 512^3$ full-box simulations: L5n512 (empty squares), L25n512 (diamonds), L50n512 (circles), L650n512 (triangles), L900n512 (pentagons). The lines are the fits to the filled in points; results from L5n512 are affected by the small size of the box and therefore were removed from the fits and are shown using empty squares. The lines become dashed where the fit is extrapolated.

The previous discussion makes clear, a single outlying particle can strongly alter the cuboid Lagrange volume; as such, it is important to ensure that the Lagrange volume is converged. Figure 3 shows the ratio between the Lagrange volume of identical halos as measured at two resolutions – defined using only halos with less
than a 5% variation in mass and position at $z = 0$ between the two resolution runs. We see that, for both $V$ and $V_{_{CH}}$, the low resolution simulation always underestimates the Lagrange volume, even for halos with more than $10^9$ particles. Extrapolating the results of Figure 5, we find that there must be at least $5 \times 10^9 - 10^6$ particles in the halo, in a "low resolution" simulation, to have a convergent Lagrange volume. The implications of this result are non-trivial: in order to converge upon the Lagrange volume of a Milky Way-size halo ($10^{12}M_{\odot}$) in a $L_{box} = 50h^{-1}$ Mpc full-box simulation, more than 2048 particles would be required for a confident measure. This is well beyond what is usually done for the first-step of a zoom simulation.

This result may initially be quite surprising. It seems unlikely that, for halos with more than $10^9$ particles in the low resolution run, the Lagrange volume is so poorly defined while all other halo properties are in excellent agreement with higher resolution simulations. Again, this can be understood by considering the increasingly irregular shape of the Lagrange volume at increasingly higher mass resolutions. Regardless of the method for determining the Lagrange volume, it ultimately depends only on a few particles — in the case of the cuboid Lagrange volume, for example, a maximum of six. As the number of particles (resolution) increases, we better sample the Lagrange volume, selecting particles from farther away that migrate into the halo. The contribution of these particles to the halo mass is insignificant, so this and other halo properties are not strongly affected, but, as we show here, it prevents full convergence of the Lagrange volume and can be important for hydrodynamic runs. We have also checked that this effect is not just due to two-body scattering effects. As the mass goes down and the force resolution goes up, it could become increasingly easier for particles to scatter far and end up somewhere close to the halo. To do this we compare the Lagrange volumes obtained from one of our standard runs ($L_{50n512}$) and an identical one in which we just used a higher value of gravitational softening that matched the one used in the next lower resolution level (i.e., $L_{50n256}$). We find that there are not significant variations or trends in the ratio of the Lagrange volumes between these runs.

These results motivate selecting a Lagrange volume for a multimass simulation using $R_{_{tb}} > R_{_{v}}$. Figure 6 shows the cuboid Lagrange volume for traceback radii between $R_{_{v}}$ and $8 R_{_{v}}$, normalized by the virial volume, as a function of $R_{_{tb}}$. We first note that the Lagrange volume grows rather slowly with $R_{_{tb}}$ — a traceback radius of $R_{_{tb}} = 5 R_{_{v}}$ increases the $z = 0$ traceback volume by 125, but the average Lagrange volume in the initial conditions increases by only $\sim 2$. Thus, the extra computational cost associated with a larger traceback radius is significantly less than one may have expected.

Figure 6 also includes results for the merger tree techniques and the adiabatic runs. In the case of the former, we find that they are generally equivalent to choosing a larger $R_{_{tb}}$ at $z = 0$. In particular, all $V_{1,m_{b}}$ and $V_{1,m_{t}}$ Lagrange volumes are enclosed by the $V^2$ choices. In the case of $V_{3,m_{b}}$ and $V_{3,m_{t}}$, there is no significant difference with $V^3$. As the merger tree analyses take a significant amount of computational effort with no obvious benefit, we will consider only the case of selecting traceback particles via a radius
defined at a single snapshot from here on. Results from the adiabatic runs are not significantly different from those computed for the same halos in a dark matter-only simulation. In the following section, we explore how contamination affects a zoom-in halo with using different $R_{tb}$.

4 CONTAMINATION

One concern with any zoom simulation is that the region of interest be free from contamination by low resolution particles throughout cosmic time. Naively, one might think this can be easily accomplished by simply choosing a Lagrange volume slightly larger than the high resolution region; however, no established approach exists in the literature — instead, each group uses its own recipe with varying degrees of success. In this section we investigate this issue systematically using several multimass simulations created with cuboid Lagrange volumes. We will later discuss the differences between this definition and that using the convex hull volume, but it is important to state here that the main conclusions of this work are valid for both definitions. We primarily focus on Milky Way-mass halos, but we also explore a few clusters and dwarf galaxy halos (see Table 2).

Figure 7 illustrates one of the main conclusions of this work: the chance for contamination increases with the level of zoom. The vertical axis shows the distance from the host halo center to the first low-resolution particles for several of our zoom simulations. Particle distances are plotted in units of the track-back radius $R_{tb}$ and shown as a function of resolution steps $\Delta_{res}$ between the full box used to calculate the Lagrange volume and the zoom simulation itself. Each $\Delta_{res}$ step corresponds to a factor of $2^3$ increase in effective particle number, such that the mass resolution improves as $m_p \rightarrow m_p/8^{\Delta_{res}}$. In ideal cases, there would be no points with $r/R_{tb} < 1$. Error bars indicate the range in distances between the first and fifth low resolution particles closest to the center of the halo. Lines connect zoom-in simulations with identical Lagrange volumes but varying the mass resolutions. Naively, one might expect the y-axis ratio to hover around 1, which would indicate that particles in multimass simulations end in roughly the same location as in full-box runs. However, we find that for a fixed Lagrange volume, low resolution particles wander closer to the center of the region of interest as the number of resolution steps increases.

4.1 Avoiding contamination

Given that the Lagrange volume of a halo can be irregularly shaped, its overall size can be affected by a few particles. As we have seen, the Lagrange volume of an individual halo will often grow when resampled at higher resolution. However, even if there is no change in the particle mass, small numerical differences from run to run can result in slightly different trajectories through time. Of course, it is possible to choose a small Lagrange volume and, purely by luck, have no contamination in the region of interest. However, this is not wise given the computational cost of many zoom simulations.

We present an empirical relation for defining a Lagrange volume that guarantees a resimulated halo is free from contamination...
within $R_v$. Specifically, for a simulation that zooms by a factor of $\Delta_{res} (m_p \rightarrow m_p/8^{\Delta_{res}})$ compared to the initial box, then the Lagrange volume can be defined using the particles within a sphere around $z = 0$ particles of radius

$$R_{th} = (1.5 \Delta_{res} + 1) \times R_v, \quad (1)$$

in order to conservatively avoid halo contamination by low res particles. As discussed above, if we increase the resolution of the zoom-in simulation we need to increase the Lagrange volume. This formula is based on the results of the $\sim 130$ multimass simulations we have studied for this work, all of which used Lagrange volumes calculated from at least 4000 particles. Figure 8 shows a set of these simulations in which empty circles indicate non-contaminated halos and filled circles signify contaminated runs. The black dashed line stands for equation $1$ (slope of $1.5$) and grey dashed line stands for the same equation but with a slope of $1$.

We stress here that Equation $1$ is derived from experimental results using full-box simulations with resolutions up to 512$^3$ particles and zoom-in simulations with effective resolutions as high as 4096$^3$. The relation is primarily driven by the correlation between the Lagrange volume and the resolution at which it is calculated. As pointed out above, we expect the Lagrange volume to converge when it is calculated from $\sim 10^5$ particles; thus, if the Lagrange volume is initially constructed from a relatively high resolution simulation, Equation $1$ may be significantly relaxed. Conversely, if the Lagrange volume is calculated with fewer than $\sim 4000$ particles, it should be greatly increased. Depending on the desired $\Delta_{res}$, it may even be computationally cheaper to use intermediate resolution runs to refine the Lagrange volume for the highest resolution (and therefore most expensive) simulation.

There is also a second-order mass dependence that we do not consider, which is directly related to the anti-correlation presented in Figure 2. However, to fully capture this effect, a much larger sample of multimass simulations is needed. Here we simply note that zoom-in simulations aimed at dwarf-sized halos require proportionally larger Lagrange volumes than those aimed at clusters.

### 4.2 Effects of Contamination

The previous section discusses in great detail how contamination can be prevented; however, one might argue that contamination on the order of a few percent does not matter. Clearly, large-scale properties (e.g. halo mass) are mostly unaffected by a small number of low resolution particles within the virial radius. However, many zoom-in simulations are performed with the goal of determining small-scale features within the halo or for gas dynamical simulations that are sensitive to the existence of contaminating massive particles. Here we explore whether even a small number of contaminating low resolution particles can alter halo properties.

We have found that for dark matter-only simulations, no halo properties (shape, concentration, spin, etc.) show significant differences, provided the contamination remains less than 2% of the mass within the virial radius. For adiabatic multimass simulations, however, we find that contamination has a significant effect on the gas properties of the host halo. As an example, we plot the gas and dark matter density profiles of a specific halo (drawn from the L50m512-adiabatic box) in Figure 8, lower panels show the deviation between contaminated and uncontaminated zoom-in runs, where $\Delta_{res} = 2$. The three multimass simulations are identical aside from their Lagrange volumes. We find that when low resolution particles are present within $R_v$, the gas density profile is significantly altered from the uncontaminated case. Such variation is evident in all contaminated adiabatic zoom-in simulations that we have run, and we suggest therefore that contaminated halos will exhibit systematically lower gas densities. Additionally, the contamination within a halo correlates with the total gas mass within $R_v$, and therefore the baryon fraction: 5% contamination results in a baryon fraction 5% smaller than that measured in an uncontaminated halo. We also note that low resolution dark matter particles within $R_v$ tend to act as sinks for gas particles, leading to artificial substructure formation. This effect is small but appreciable in our adiabatic test runs; including cooling and star formation should act to exacerbate such numerical artifacts.

Finally, we investigate possible edge effects in our adiabatic simulations. Typically only the high resolution region in a baryonic Lagrangian zoom-in simulation is populated with gas particles; therefore, an artificial boundary condition is created at the edge of the high resolution region wherein the gas density instantaneously drops to zero. Eulerian zoom-in simulations use a low resolution grid outside of the high resolution region and therefore have no sharp edge. Though one may worry that such an edge creates unphysical effects in the gas properties, we find no relevant edge effects as the profiles only diverge at well beyond low resolution particles. Any edge effects, therefore, are completely overwhelmed by errors introduced by contamination.

### 5 SELECTING A HALO FOR RESIMULATION

As Figures 2 through 6 show, the scatter in the Lagrange volume of a halo at fixed mass is large, regardless of the volume definition. Larger Lagrange volumes are necessarily more expensive computationally; choosing a halo with a small Lagrange volume is the easiest way to reduce the required memory and CPU time. However, choosing halos with small Lagrange volumes may bias one towards specific halo properties. In this section, we explore the correlations between Lagrange volume and other halo properties. We focus on...
the Lagrange volume as defined using $R_{tb} = 3R_v$, but all trends here can be extrapolated to different traceback radii.

5.1 Lagrange Volume Correlations

We study the relationships between the different definitions of Lagrange volume and other halo properties mentioned in Section 3. We have created a sample of well-resolved host halos that spans 8 decades in halo mass by combining all full-box simulations with $N_p = 512^3$ (L5n512, L25n512, L50n512, L650n512 and L900n512). As discussed in Section 3, $V$ correlates strongly with $M_v$; however, the normalized Lagrange volume is anti-correlated with the virial mass — more massive halos have smaller normalized Lagrange volumes (Figure 8, Spearman correlation coefficient of $-0.6$). Consequently, the normalized Lagrange volume exhibits the same correlations with other halo parameters as the virial mass (Jeeon-Daniel et al. 2011).

However, when initializing a multismass simulation, one typically has a specific halo mass in mind. We thus investigate whether the Lagrange volume correlates with any other halo properties (See Table 3) at fixed virial mass. Remarkably, aside from the expected (near-perfect) correlation with $V_{C,H}$, the only correlations we find with $V$ (with Spearman coefficient greater than 0.3) involve environmental measures. This is illustrated in Figure 10, which shows the correlation between $V$ and other halo properties for a narrow mass range of halos ($10^{11.8} < M_v/M_\odot < 10^{12.2}$), with the Spearman correlation listed in each panel. In particular we show relationships between $V$ and concentration ($V_{max}/V_v$), formation time, spin, number of subhalos, sphericity, environment (as quantified by $D_{1.0.1}$ and $N_{neigh}^3$ — see Section 2.1), and the convex hull Lagrange volume.

The correlation between Lagrange volume and environment is notable. One would expect that the highest density regions would show larger $V$ values, given that these regions are likely more chaotic at high redshift, resulting in a greater likelihood that initially outlying particles will accrete onto the halo. Interestingly, however, at almost all environments (except for the densest of all) there remains quite a spread in halo Lagrange volumes. One implication is that choosing an isolated halo for resimulation does not guarantee a small Lagrange volume.

In summary, we find that at fixed virial mass, a halo’s formation history, structure, and dynamic state is largely independent of its Lagrange volume. We would expect to see correlations with the halo’s formation history and thus other halo properties if we instead examined the density fluctuations of matter in the Lagrange volume (Lee et al. 2009; Ludlow & Porciani 2011), but as we have noted before, the size of the Lagrange volume is defined by the small subset of particles that travel the farthest to join the halo. These particles contribute very little to the total mass and thus, for a fixed mass, the Lagrange volume of a halo is unrelated to the exact nature of that halo’s collapse.

5.2 Choosing Halos to Resimulate

We have just established that Lagrange volume is uncorrelated with internal halo properties. There is some correlation with environment, but the scatter about this correlation is large. The implication is that the most efficient targets for zoom simulations are those
with the smallest normalized Lagrange volumes — such halos are the cheapest to run and sample the parameter space of dark matter halos in a largely unbiased way. This point is illustrated in Figure 11. The left panel shows the normalized cuboid Lagrange volume as a function of virial mass, with white triangles indicating halos with \(N\) in the smallest 5% of the distribution. The two other panels illustrate that this subsample of small-\(N\) halos is statistically indistinguishable from the full sample, with environment the lone exception. The small-volume subsample is clearly biased towards low density environments, as expected, though some small-volume halos do exist in high density environments.

By selecting systems with small Lagrange volume for resimulation, one can dramatically reduce the computational requirements. The Lagrange volume of an average halo is typically \(8 \rightarrow 10\) larger than those in the 5% tail, and this directly translates to unnecessary computational complexity. Provided that the environment is not an important variable to the problem of interest, small-volume halos are ideal candidates for resimulation. However, for problems that are closely tied to the environment, one must consider more elaborate methods to build a fair sample of halos while still considering the Lagrange volume, in order to reduce the cost of the total sample.

5.3 Are There Pathological Halos to Avoid?

In order to study the reliability of zoom-in simulations, specifically the effects of contamination on dark matter halos, we first needed to understand the stability of halo properties in full-box simulations. Such properties could, in principle, be sensitive to the resolution \((N_p)\), the initial redshift, or just the specific transfer function used. We therefore re-ran several of our full-box simulations, varying these parameters. We obtain a reasonable estimate for the resultant scatter by comparing the properties of identical halos in these different runs, which we use to determine the importance of contamination, Lagrange volume definitions, and sizes of buffer zones in the multimass simulations. Figure 12, which we discuss throughout this section, illustrates a few of these tests.

Trajectories within virialized systems are chaotic. Small differences existing at any time grow quickly. These differences originate in the dynamical instability of particle trajectories in high-density regions (e.g., Knebe et al. 2000; Valluri et al. 2007); in particular, as was first noted by Miller (1964), the N-body problem is chaotic in the sense that the trajectory of the 6N-dimensional phase-space coordinate of the system exhibits exponential sensitivity under small changes in the initial conditions. This exponential instability, referred to in the literature as the Miller instability or “minichaos”, has been investigated extensively in several studies (Kandrup & Smith 1991; Goodman et al. 1993). Therefore changes in the initial conditions, even for a fixed resolution (e.g. using a different \(z_{ini}\)), will cause the particles to experience slightly different trajectories, which produces an artificial scatter in the halo properties.

Most halo properties exhibit a relatively small scatter as the resolution and \(z_{ini}\) vary, with median values in the ratio of structural quantities of approximately unity, as illustrated in the upper panel of Figure 12. However, properties that are strongly affected by substructure, and therefore the specific positions of the particles, such as the spin or the halo shape (lower panel of Figure 12), while still convergent, show a significantly larger scatter about the median. The scatter in these properties can be reduced significantly if particles bound to substructures are eliminated from the calculation (e.g., Vera-Ciro et al. 2011).

For full-box halo samples, one can safely conclude that halo properties do not vary in any systematic sense. Nevertheless, at least two other interesting questions exist for multimass simulations. First, do the variations between simulations correlate with any specific halo property? That is, are there halos that are most likely to have a large scatter between runs be identified beforehand by any \(z = 0\) property? Second, do those halos that exhibit a large dispersion between two runs also exhibit substantial scatter when simulated with a third \(z_{ini}\) or at a different resolution? Specifically, do certain halos tend to have a larger dispersion between their parameters? If either is the case, one would like to identify such halos and avoid them when selecting halos for resimulation. Our analysis of the full-box simulations presented here indicates that neither is true: we find that no \(z = 0\) property is correlated with a larger dispersion in halo parameters after a resimulation at different resolution or with a different \(z_{ini}\). We also found that when comparing halos between three different re-simulations there are no specific halos that tend to have a bigger dispersion. Our results are consistent with the hypothesis that variations are driven by the number of particles that begin with high density fluctuations (Knebe et al. 2009).

6 CREATING MULTIMASS INITIAL CONDITIONS

6.1 Starting Redshift

The initial redshift \(z_{ini}\) is a key choice for any cosmological simulation. Usually, \(z_{ini}\) is chosen such that the resulting variance of
the discrete density field \(\sigma_L\), is between 0.1 and 0.2 (see Knebe et al. 2009 for a detailed discussion). Though one might expect that a higher \(z_{ini}\) is necessarily better, cosmological codes can introduce unnecessary small-scale noise during the linear regime. If the value of \(z_{ini}\) is too low, however, non-linear evolution will be compromised.

Improvements in the techniques used to calculate perturbations have reduced errors in the linear regime and allowed for lower values of \(z_{ini}\) (see, for example Jenkins 2010); some groups (e.g. Lukic et al. 2007; Prunet et al. 2008; Knebe et al. 2009; Jenkins 2010; Reed et al. 2013) have worked to quantify the effect of \(z_{ini}\) on the final results of cosmological simulations. For multimass resimulations, \(z_{ini}\) is typically identical to that of the full-box simulation, as the calculated Lagrange volume is only valid at that redshift; the Miller instability discourages one from using different initial redshifts, even if both runs are at the same resolution.

Our approach for selecting \(z_{ini}\) is to compute the values of \(z_{ini}^{0.1}\) and \(z_{ini}^{0.2}\) using the full box at the lowest redshift (that used to calculate the Lagrange volume) and the highest resolution of relevance (corresponding to the zoom-in simulation). These two redshift ranges will overlap unless \(\Delta v_{sys}\) is particularly large (which is not recommended). For example, a box size of \(650h^{-1}\) Mpc with our adopted cosmology gives initial redshift values of \(z_{ini}^{0.1} = 27.81\) and \(z_{ini}^{0.2} = 13.40\) for \(N_p = 512^3\) run, and the same simulation with an effective resolution of \(2048^3\) corresponds to \(z_{ini}^{0.1} = 46.10 - z_{ini}^{0.2} = 22.55\). Therefore an initial redshift between 27.81 and 22.55 would be recommended for both simulations.

We have re-run several full-box simulations with different initial redshift to ensure agreement with results in the literature. In Figure 13, we plot the evolution of the Lagrange volume of a L50n256 simulation for halos at \(z = 0\) (black points) along with the \(z_{ini}\) Lagrange volume computed for two other identical L50n256 simulations but with a lower \(z_{ini}\) (red triangles). The calculated Lagrange volume is largely independent of the initial redshift and, in fact, \(V\) evolves very slowly for redshifts greater than the (mass dependent) turn-around redshift; therefore, all of our results concerning the Lagrange volume are valid for any \(z_{ini}\) greater than the turn-around redshift. In this Figure we also plot the \(z_{ini}\) Lagrange volume computed for three identical L50n256 simulations runs using ENZO (blue circles) which show a perfect agreement with GADGET-2 results.

6.2 Lower resolution regions

Initial conditions for multimass simulations typically surround the high resolution region with shells, or buffer volumes, of progressively lower resolution particles, with the number of shells varying
from 0 to $\Delta_{res} - 1$. Again, many groups have developed a methodology based on trial and error, but our literature search revealed no quantitative analysis. We explore how the properties of these buffer volumes, specifically the size, number, and steps in resolution between them, affect uncontaminated halos at $z = 0$.

Our results indicate that using a step in resolution larger than one between successive shells, i.e. a factor of $8^2$ or greater increase in particle mass, strongly alters the properties of the halo at $z = 0$. It is unclear whether the variance in the high resolution region is due to the mixing of dramatically different particle masses, which leads to poorly resolved structures in the buffer volumes, or because the low mass particles near the edge of the high resolution volume have their trajectories too strongly affected by the high mass particles nearby, but the effects also propagate into uncontaminated regions. Therefore, we recommend using a buffer zone for each step in resolution (factor eight in mass) between the largest and smallest particle masses.

Similarly, the host halo will not be properly converged if the buffer volumes are too small; however, increasing the size beyond this minimum value does not effect the final properties in any way but increases the CPU and memory cost of the simulation. As is the default in MUSIC, we created multimass initial conditions such that the width of a given level, $w_i$, sets the width of next lower resolution grid level, $w_{i-1}$. In units of the grid size of the highest resolution grid in the simulation (i.e. the Lagrange volume),

$$w_{i-1} = w_i / 2 + c,$$

where $c$ is a constant added to guarantee that each shell has a minimum width; that is, the second highest resolution region is half as wide ($+c$) as the Lagrange volume in the initial conditions. Test multimass simulations show that variations in $c$ do not alter the final halo, provided that $c \geq 12$. For users of MUSIC, we note that this corresponds to padding $\geq 6$, as $c = 2 \times$padding.

Finally, using the recipe described above, we varied the lowest resolution level that comprises the majority of the multimass box, which fixes the number of buffer zones. As expected, we found that there is generally a minimum for this resolution: it the large-scale tidal must be high enough to capture forces relevant to the targeted halo. The multimass simulations studied in this work indicate that the host halo is converged provided the lowest effective resolution is at least $128^3$. Again, setting this resolution higher results in unnecessary, and significant, memory and CPU cost.

### 6.3 Other Definitions of Lagrange Volume

We also present results of multimass simulations using different Lagrange volume definitions. Though we have not run as many simulations for these cases, there are a number of interesting results. We have initialized multimass simulations using the convex hull Lagrange volumes instead of the cuboid by creating full-box initial conditions at all required resolutions, selecting from the highest resolution box all particles that lie within $V_{CH}$, then filling the remainder of the volume with particles from the low resolution initial conditions. We ran twelve GADGET-2 simulations of halos with mass $M_0 \simeq 10^{12} M_\odot$. These are not listed in Table 2. For a fixed $R_{tb}$, this technique increases the likelihood of contamination compared with the cuboid approach. It therefore requires larger values of $R_{tb}$ to avoid contamination. However, even taking this into account, we found that using this technique, one may typically resolve uncontaminated halos with 40 – 50% fewer particles than with the cuboid Lagrange volume. To ensure uncontaminated halos using this technique (in parallel with Equation[1]), we recommend a convex hull Lagrange volume defined by

$$R_{tb} = (1.5 \times \Delta_{res} + 7) \times R_c. \quad (2)$$

MUSIC has recently added support to generate initial conditions using also the minimum ellipsoid Lagrange volume ($V_{ME}$). Running a full set of simulations using this Lagrange volume definition is beyond the scope of this paper. However the convex hull results (Eq.[2]) will ensure uncontaminated halos for this Lagrange volume definition as well.

### 6.4 Eulerian Codes

We have also run full box and multimass simulations using ENZO. We already showed that, for the same initial conditions, Lagrange volumes obtained using ENZO or GADGET-2 are very similar. In the case of multimass simulations, the similar effects of low resolution contamination were also found with ENZO; that is, for a fixed Lagrange volume, low resolution particles get closer to the center of the halo as the level of zoom increases. Therefore the difference in resolution between that used to calculate the Lagrange volume and that used in the high resolution region of the multimass simulation has to be taken into account. We also found that due to the adaptive mesh refinement algorithm (AMR) in Eulerian codes, it is recommendable to include the region in which the halo is located along redshift when calculating the Lagrange volume. We studied this property and found that the distance between the center of mass at $z_{ini}$ and $z = 0$ only shows a weak correlation with the environmental parameter, $N_{3d}^{neigh}$. Therefore, this parameter can also be used to pick halos for zoom-in simulations without significantly biasing the final sample. Finally, in the case of ENZO we find much more scatter between similar zoom-in simulations and therefore do not want to give an explicit recommendation concerning the size of the Lagrange volume. The study of a higher statistical ENZO sample is beyond the scope of this paper.

### 7 CONCLUSION AND DISCUSSION

We have run a large number of cosmological full-box and zoom-in simulations of various sizes, particle number, and initial redshifts. By analyzing the properties of the halos in these simulations, we studied the methodology behind the multimass technique and have arrived at a number of conclusions and recommendations regarding the Lagrange volume, contamination, and how to best perform zoom simulations.

In the first part of this paper we presented a detailed study of the high-redshift Lagrange volume associated with the makeup of $z = 0$ halos, including a variety of definitions for calculating the Lagrange volume itself. Importantly, we have found that the Lagrange volume of any given halo converges very slowly as the mass resolution of the simulation increases, and that there can be catastrophic errors if fewer than $\sim 500$ particles are used. Though halo virial masses are well-converged, Lagrange volume regions can be quite irregular, with overall volumes governed by the furthest few particles. Extrapolating the results of Figure 4, we find that approximately $\sim 10^5$ particles are required in a low resolution halo in order to have a confidently converged Lagrange volume. This result is related to a particularly striking finding (Figure 6): that the chance for low-resolution particle contamination increase steadily with the level of zoom. Equation[1] provides a formula for choosing the Lagrange volume conservatively to avoid halo contamination...
based on the level of refinement between the parent simulation and
the ultimate resimulation volume. This formula was derived using
\( \sim 130 \) zoom-in simulations that focused on objects ranging from
dwarf galaxy halos to clusters, and is applicable for Lagrangian
codes.

In studying the general properties of dark-matter only halos
with 1-2% contamination by mass, we do not find significant varia-
tion in the fundamental halo properties (formation time, concentra-
tion, shape, etc.). However, we do find important divergences be-
tween the halo properties of contaminated and non-contaminated
halos in adiabatic gas multimass runs, such as spurious baryonic
structure formation, and therefore strongly recommend the usage
of Lagrange volumes that guarantee uncontaminated halos for all
hydrodynamic simulations.

We have also performed a detailed statistical analysis between
the varying Lagrange volume definitions and other typical halo
properties using a sample of halos spanning eight orders of mag-
nitude in mass from a set of full-box simulations. After normal-
izing the Lagrange volume by the virial volume of each halo, we
find an anticorrelation between Lagrange volume and virial mass.
Smaller halos therefore have proportionally larger Lagrange vol-
umes than more massive ones. Thus, cluster zoom-in simulations
are cheaper than those focused on dwarf galaxy halos, for a fixed
number of particles within the halo. However, a fixed halo mass, the
Lagrange volume does not correlate with any internal halo property
we have considered. We find a mild correlation with large-scale en-
vironment, such that the most clustered halos have larger Lagrange
volumes.

Based on our results we provide provide the following recom-
endations for performing multi mass zoom-in simulations:

- Run the full-box simulation from which the zoom-in halos will
  be selected at the highest resolution possible compared to the ul-
timate zoom-in simulation. This will allow for more accurate values
  of the Lagrange volumes to build initial conditions and decrease
  the chances for contamination.

- Select the initial redshift, \( z_{ini} \), for the full-box simulation
  based on the highest resolution zoom-in simulation planned. Al-
though we showed that the Lagrange volume does not change sig-
nificantly at high redshift, this will reduce the mini-chaos effect in
the simulation and help the comparison with the full box results.

- Do not pick halos for zoom-in simulations with fewer than 500
  particles in the full-box run – the Lagrange volume will dramati-
cally increase with increasing resolution for such poorly resolved
halos.

- Avoid contamination in the final halo by accounting for the
  difference in resolution between the full-box and the zoom-in simu-
lation when selecting the Lagrange volume (see Equations\[1\] and\[2\]).

- For a given halo mass, select those halos with the smallest
  Lagrange volumes for zoom-in simulations, in order to minimize
  CPU and memory cost.

- Create a buffer volume at each resolution step between the
  high resolution region and the lowest resolution that comprises
  the majority of the box; however, the size of these volumes and the
  lowest effective resolution should both be minimized to again eliminate
  unnecessary CPU and memory usage. The simulations analyzed in
this work indicate that using 128\(^3\) for the latter effectively captures
the large-scale tidal forces; we elaborate on the minimum width in
Section\[6\], but note for users of MUSIC that our results require
\( \text{padding} \geq 6 \).

- In the case of the most expensive multimass hydrodynamical
  simulations it is highly advisable to first run the collisionless version
with the same Lagrange volume to guarantee non-contaminated ha-
las.

**ACKNOWLEDGMENTS**

This work used computational resources granted by NASA Ad-
vanced Supercomputing (NAS) Division, Nasa Center for Climate
Simulation, Teragrid and by the Extreme Science and Engineering
Discovery Environment (XSEDE), which is supported by National
Science Foundation grant number OCI-1053575. We also made use
of resources at the City University of New York High Performance
Computing Center which is supported, in part, under National Sci-
cence Foundation Grants CNS-0958379 and CNS-0855217 and the
GreenPlanet cluster at UCI. JO, SGK, JSB, and MR were supported
by NSF grant AST-1009999 and NASA grant NNX09AG01G. JO
also thanks the financial support of the Fulbright/MICINN Pro-
gram.

**REFERENCES**

Allgood B., Flores R. A., Primack J. R., Kravtsov A. V., Wechsler
R. H., Faltenbacher A., Bullock J. S., 2006, MNRAS, 367, 1781
Barber C. B., Dobkin D. P., Huhdanpaa H., 1996, ACM Trans.
Math. Softw., 22, 469
Bertschinger E., 2001, ApJS, 137, 1
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A.,
Porciani C., Primack J. R., 2001, ApJ, 555, 240
Feldmann R., Carollo C. M., Mayer L., Renzini A., Lake G.,
Quinn T., Stinson G. S., Yepes G., 2010, ApJ, 709, 218
Frenk C. S., Evraud A. E., White S. D. M., Summers F. J., 1996,
ApJ, 472, 460
Garrison-Kimmel S., Boylan-Kolchin M., Bullock J., Lee K.,
2013, in prep.
Goodman J., Heggie D. C., Hut P., 1993, ApJ, 415, 715
Haas M. R., Schaye J., Jeeson-Daniel A., 2012, MNRAS, 419,
2133
Hahn O., Abel T., 2011, MNRAS, 415, 2101
Hahn O., Teyssier R., Carollo C. M., 2010, MNRAS, 405, 274
Howlett C., Lewis A., Hall A., Challinor A., 2012, Journal of Cos-
ology and Astroparticle Physics, 4, 27
Jeeson-Daniel A., Dalla Vecchia C., Haas M. R., Schaye J.,
2011, MNRAS, 415, L69
Jenkins A., 2010, MNRAS, 403, 1859
Joyce M., Marcos B., Baertschiger T., 2009, MNRAS, 394, 751
Kandrup H. E., Smith Jr. H., 1991, ApJ, 374, 255
Katz N., White S. D. M., 1993, ApJ, 412, 455
Khachiyan L. G., 1979, Doklady Akademiia Nauk SSSR, 244,
10931096
Klypin A., Kravtsov A. V., Bullock J. S., Primack J. R., 2001,
ApJ, 554, 903
Knebe A., Kravtsov A. V., Gottlöber S., Klypin A. A., 2000,
MNRAS, 317, 630
Knebe A., Wagner C., Knollmann S., Dieckershoff T., Krause F.,
2009, ApJ, 698, 266
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Komatsu E., Smith K. M., Dunkley J., et al., 2011, ApJS, 192, 18
Lee J., Hahn O., Porciani C., 2009, ApJ, 707, 761
Lewis A., Challinor A., Lasenby A., 2000, ApJ, 538, 473
Ludlow A. D., Porciani C., 2011, MNRAS, 413, 1961
