Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **440,** 2115–2135 (2014) doi:10.1093/mnras/stu390

N-body dark matter haloes with simple hierarchical histories

Lilian Jiang,* John C. Helly, Shaun Cole and Carlos S. Frenk

Institute for Computational Cosmology, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK

Accepted 2014 February 24. Received 2014 February 10; in original form 2013 November 26

ABSTRACT

We present a new algorithm which groups the subhaloes found in cosmological *N*-body simulations by structure finders such as SUBFIND into dark matter haloes whose formation histories are strictly hierarchical. One advantage of these 'Dhaloes' over the commonly used friends-of-friends (FoF) haloes is that they retain their individual identity in the cases when FoF haloes are artificially merged by tenuous bridges of particles or by an overlap of their outer diffuse haloes. Dhaloes are thus well suited for modelling galaxy formation and their merger trees form the basis of the Durham semi-analytic galaxy formation model, GALFORM. Applying the Dhalo construction to the Λ cold dark matter Millennium II Simulation, we find that approximately 90 per cent of Dhaloes have a one-to-one, bijective match with a corresponding FoF halo. The remaining 10 per cent are typically secondary components of large FoF haloes. Although the mass functions of both types of haloes are similar, the mass of Dhaloes correlates much more tightly with the virial mass, M_{200}, than FoF haloes. Approximately 80 per cent of FoF and bijective and non-bijective Dhaloes are relaxed according to standard criteria. For these relaxed haloes, all three types have similar concentration–M_{200} relations and, at fixed mass, the concentration distributions are described accurately by log-normal distributions.

Key words: methods: numerical – galaxies: haloes – cosmology: theory – dark matter.

1 INTRODUCTION

In hierarchical dark matter-dominated cosmologies, such as standard Λ cold dark matter (CDM), galaxy formation is believed to be intimately linked to the formation and evolution of dark matter haloes. Baryonic gas falls into dark matter haloes, cools and settles into centrifugally supported star-forming discs (Binney 1977; Rees & Ostriker 1977; White & Rees 1978; White & Frenk 1991; Kauffmann & White 1993; Cole et al. 1994; Somerville & Primack 1999; Benson et al. 2003). Thus, the evolution of the galaxy population is driven by the evolution of the population of dark matter haloes which grow hierarchically via mergers and accretion. Therefore, to model galaxy formation, one must first have an accurate model of the evolution of dark matter haloes.

The formation and evolution of dark matter haloes from cosmological initial conditions in large representative volumes can now be routinely and reliably simulated using a variety of *N*-body codes (e.g. Springel 2005). In contrast, simulations of the evolution of the baryonic component are much more uncertain with gross properties depending on the details of uncertain subgrid physics as well as on the limitations of numerical hydrodynamics (Schaye et al. 2010; Creasey et al. 2011). Hence, a useful and complementary approach is semi-analytic galaxy formation (e.g. Cole 1991; Lacey & Silk 1991; White & Frenk 1991; Kauffmann & White 1993; Cole et al. 1994, 2000; Somerville & Primack 1999; Bower et al. 2006; Somerville et al. 2008; Benson & Bower 2010) in which one starts with the framework provided by the dark matter halo evolution and uses analytic models to follow the processes of galaxy formation that occur within these haloes. The key starting point for this approach is halo merger trees which quantify the hierarchical growth of individual dark matter haloes.

In ΛCDM, the first self-bound objects to form are haloes with masses of around an Earth mass corresponding to the small-scale thermal cut-off in the CDM power spectrum (Green, Hofmann & Schwarz 2004). In a cosmological *N*-body simulation, the mass scale of the first generation of haloes is instead set by the mass resolution of the simulation. Subsequent generations of haloes form by mergers of earlier generations of haloes plus some smoothly accreted material. The merging process does not produce a completely relaxed smooth halo, and the remnants of the earlier generation of haloes are often detectable as self-bound substructures (subhaloes) within the new halo. Thus, it is important to distinguish between haloes and the subhaloes that they contain which are the remnants of early generations of now merged haloes. A variety of algorithms which can identify these subhaloes in *N*-body simulations have been devised (Onions et al. 2012). These substructure finders are capable of detecting arbitrary levels of nested subhaloes within subhaloes and in most cases also identify the background mass distribution in a halo as a subhalo. In this work, we refer to all of the groups identified by such substructure finders as 'subhaloes' and merger trees constructed by identifying a descendant for each subhalo as

* E-mail: lilian.jiang@durham.ac.uk

'subhalo merger trees'. Srisawat et al. (2013) compare a range of methods for the production of subhalo merger trees. The algorithm we use to determine subhalo descendants in this paper is included in the comparison under the name D-TREES.

To construct the halo merger trees needed by semi-analytic models, it is not sufficient to just track subhaloes between simulation outputs (e.g. by tracking their constituent particles), but one also needs to identify their host haloes. For instance, when a galaxy cluster forms, it is normally assumed that while the galaxies remain in their individual subhaloes the diffuse gas surrounding them and gas blown out of the galaxies by supernova (SN)-driven winds are not retained by the individual subhaloes but instead join the common intracluster medium of the surrounding halo of the galaxy cluster. Another issue that has to be addressed when building merger trees for use in galaxy formation models is that structure formation for the collisionless material in *N*-body simulations is not strictly hierarchical. Hence, occasionally when two haloes merge, the subhalo resulting from the smaller progenitor can pass straight through the main halo and escape to beyond its virial radius. For the galaxy formation process to be followed, it is necessary to retain the association between these two separated subhaloes so that an appropriate physical model can be applied to their diffuse collisional gas which would not have separated after the merger. Merger trees that are useful for galaxy formation modelling have to take account of these considerations (Knebe et al. 2013). The Dhalo algorithm which we present produces a set of haloes which is strictly hierarchical in the sense that once a subhalo becomes a component of a Dhalo, it never subsequently demerges.

It is now quite common for semi-analytic models to use halo merger trees extracted directly from *N*-body simulations (Springel et al. 2001; Hatton et al. 2003; Helly et al. 2003; Bower et al. 2006; Koposov et al. 2009; Muñoz et al. 2009; Busha et al. 2010; Macciò et al. 2010; Guo et al. 2011). There are many choices to be made both in defining the halo catalogues and in constructing the links between haloes at different times. Knebe et al. (2011, 2013) have found significant differences in even the most basic properties (e.g. the halo mass function) of halo catalogues constructed with different group finding codes. Additionally, these halo catalogues can often be modified by the procedure of constructing the merger trees as some of the algorithms break up or merge haloes together in order to achieve a more consistent membership over time (Helly et al. 2003; Behroozi et al. 2013). So, for example, even if one starts with standard friends-of-friends (FoF) groups (Davis et al. 1985), the process of building the merger trees can alter the abundance and properties of the haloes.

Semi-analytic models such as GALFORM have the option of using information extracted directly from an *N*-body simulation or using Monte Carlo methods (see Jiang & van den Bosch 2014 for a comparison of different algorithms) which make use of statistical descriptions of *N*-body results such as analytic halo mass functions (e.g. Sheth & Tormen 1999; Jenkins et al. 2001; White 2001; Evrard et al. 2002; Linder & Jenkins 2003; Reed et al. 2003, 2007; Lokas, Bode & Hoffmann 2004; Heitmann et al. 2006; Warren et al. 2006; Tinker et al. 2008; Boylan-Kolchin et al. 2009; Lukic et al. 2009; Crocce et al. 2010; Bhattacharya et al. 2011; Courtin et al. 2011; Watson et al. 2013) and models for the distribution of the concentrations of halo mass profiles (e.g. Navarro, Frenk & White 1995, 1996; Bullock et al. 2001; Eke, Navarro & Steinmetz 2001; Macciò, Dutton & van den Bosch 2008). These statistical descriptions are often based on the abundance and properties of FoF haloes and so may not be directly applicable to the catalogues of haloes that result from the application of a specific merger tree algorithm. The internal structure of the dark matter haloes strongly influences galaxy formation models. Often the gas density profiles within dark matter haloes are assumed to be related to the dark matter profile, e.g. through hydrostatic equilibrium and these influence the rate at which gas cools on to the central galaxy. In addition, the central potential of the dark matter halo affects the size and circular velocity of the central galaxy which in turn can have a strong effect on the expulsion of gas from the galaxy via SN feedback. Hence, for semi-analytic galaxy formation modelling, it is important to adopt models of the individual haloes that are consistent with the haloes that appear in the merger trees used by the semi-analytic model.

In this paper, we present a detailed description of the latest *N*-body merger tree algorithm that has been developed for use with the semi-analytic code GALFORM. The algorithm is an improvement over the earlier version, described in Merson et al. (2013), which was run on the Millennium Simulation (Springel 2005) and widely exploited in a range of applications (Bower et al. 2006; Font et al. 2008; Kim et al. 2011; Merson et al. 2013). The resulting differences between the two algorithms are very small when applied to relatively low resolution simulations such as the Millennium, but the improvements in the new algorithm do a better job of tracking halo descendants in high-resolution simulations such as the Millennium II (MSII; Boylan-Kolchin et al. 2009) and Aquarius simulations (Springel et al. 2008). The starting point for our merger trees is FoF haloes that are decomposed into subhaloes, distinct self-bound structures, by the substructure finder, SUBFIND (Springel et al. 2001). Subhaloes are tracked between output times and agglomerated into a new set of haloes, dubbed Dhaloes, that have consistent membership over time in the sense that once a subhalo is accreted by a Dhalo it never demerges. In this process, we also split some FoF haloes into two or more Dhaloes when SUBFIND substructures are well separated and only linked into a single FoF halo by bridges of low-density material.

Our paper is structured as follows. In Section 2, we briefly outline the new merger tree algorithm (full details are given in Appendix A) and its application to the MSII Simulation. In Section 3, we compare and contrast the properties of the resulting Dhaloes with the more commonly used FoF haloes (Davis et al. 1985). We show specific rare examples where Dhaloes and their matched FoF counterparts exhibit gross differences either one FoF halo being decomposed into several Dhaloes or vice versa. We also examine the distribution of mass ratios for matching Dhalo and FoF pairs. Then in Section 4 we compare statistical properties of halo populations including halo mass functions and their concentration–mass relation. We conclude in Section 5.

2 HALO CATALOGUES

Immediately below, Section 2.1, we summarize the specification of the MSII Simulation which we use to test and illustrate the application of our merger tree algorithm. We then give a brief outline of the construction of the merger trees and their constituent haloes with the complete specification detailed in Appendix A.

2.1 The MSII Simulation

The MSII Simulation[1] (Boylan-Kolchin et al. 2009) was carried out with the GADGET3 *N*-body code, which uses a 'TreePM' method

[1] The MSII Simulation data will be available from a Structured Query Language relational data base (Lemson 2006) that can be accessed at http://galaxy-catalogue.dur.ac.uk:8080/Millennium.

to calculate gravitational forces. The MSII is a cosmological simulation of the standard ΛCDM cosmology in a periodic box of side $L_{\rm box} = 100\ h^{-1}$ Mpc containing $N = 2160^3$ particles of mass $6.95 \times 10^6\ h^{-1}\ {\rm M}_\odot$. The cosmological parameters for the MSII are $\Omega_{\rm m} = 0.25$, $\Omega_{\rm b} = 0.045$, $h = 0.73$, $\Omega_\Lambda = 0.75$, $n = 1$ and $\sigma_8 = 0.9$. Here $\Omega_{\rm m}$ denotes the total matter density in units of the critical density, $\rho_{\rm crit} = 3H_0^2/(8\pi G)$. $\Omega_{\rm b}$ and Ω_Λ denote the densities of baryons and dark energy at the present day in units of the critical density. The Hubble constant is $H_0 = 100\,h\,{\rm km\,s^{-1}\ Mpc^{-1}}$, n is the primordial spectral index and σ_8 is the rms density fluctuation within a sphere of radius $8\,h^{-1}$Mpc extrapolated to $z = 0$ using linear theory. These cosmological parameters are consistent with a combined analysis of the 2dFGRS (Colless et al. 2001; Percival et al. 2001) and first year *Wilkinson Microwave Anisotropy Probe* data (Spergel et al. 2003; Sanchez et al. 2006).

2.2 Building merger trees

The first step in building our merger trees is the construction of catalogues of both FoF haloes (Davis et al. 1985) and their internal self-bound substructures,[2] subhaloes, as identified by SUBFIND (Springel et al. 2001). The second step is to build SUBFIND merger trees by tracking particles between output snapshots to determine the descendant of each subhalo. Occasionally, SUBFIND fails to find a substructure as it transits through the core of a larger halo. To avoid this resulting in the premature merging of substructures, we have developed an algorithm (Appendix A2) that looks several snapshots ahead to robustly link progenitor and descendant subhaloes. A similar approach was adopted by Behroozi et al. (2013) to construct self-consistent merger trees for the Bolshoi simulations (Klypin, Trujillo-Gomez & Primack 2011). The third step is to partition these SUBFIND merger trees into discrete branches. A new branch begins whenever a new subhalo forms and continues for as long as the subhalo exists in the simulation. When a merger occurs, we decide which of the progenitor subhaloes survives the merger by determining which progenitor contributed the most bound part of the descendant (see Appendix A2.1). The branch corresponding to this progenitor continues, while the other progenitor's branch ends. The final step is to bundle these branches together to define the composite Dhaloes and their merger trees. Here our algorithm (described in full in Appendix A3) defines collections of subhaloes embedded hierarchically within each other as a single Dhalo, but excludes neighbouring subhaloes that may be part of the same FoF group, but are only linked in by a bridge of low-density material or subhaloes that are beginning the process of merging but have not yet lost a significant amount of mass. Subhaloes are grouped into Dhaloes in such a way that once a subhalo becomes part of a Dhalo, it remains a component of that Dhalo's descendants at all later times at which the subhalo survives, even if it is a satellite component that takes it temporarily outside the corresponding FoF halo. All of a Dhalo's subhaloes which survive at a later snapshot must belong to the same Dhalo at that snapshot. We take this to be the descendant of the Dhalo. This defines the Dhalo merger trees. The mass of a Dhalo is simply the sum of the masses of its component subhaloes.[3]

[2] Here we identify the subhaloes using the SUBFIND algorithm (Springel et al. 2001). However, this is not the only option and there is now a large literature (see Onions et al. 2012) on alternative methods of identifying self-bound structures. Some of these are highly sophisticated and use full 6D phase space information to disentangle spatial coincident subhaloes (Diemand, Kuhlen & Madau 2006; Behroozi, Wechsler & Wu 2012). As an example, we have experimented with building Dhaloes by applying the Dhalo algorithm from Appendix A3 onwards but with SUBFIND subhalo merger trees replaced by those defined by the Hierarchical Bound Tracing (HBT) algorithm of Han et al. (2012). We find that the properties of the Dhalo merger trees and the galaxies that result after they are processed by GALFORM are extremely similar.

3 COMPARISON OF FoF AND DHALOES

3.1 Bijectively matched FoF and Dhaloes

The properties of FoF haloes, especially those defined by the conventional linking length parameter of $b = 0.2$ (the linking length is defined as b times the mean inter-particle separation), are well documented in the literature (e.g. Huchra & Geller 1982; Press & Davis 1982; Einasto et al. 1984; Davis et al. 1985; Frenk et al. 1988; Lacey & Cole 1994; Summers, Davis & Evrard 1995; Audit, Teyssier & Alimi 1998; Klypin et al. 1999; Jenkins et al. 2001; Eke et al. 2004; Warren et al. 2006; Gottlöber & Yepes 2007), and such haloes are widely used as the starting point for relating the dark matter and galaxy distributions (Peacock & Smith 2000; Seljak 2000; Berlind & Weinberg 2002). Thus, as the semi-analytic model GALFORM (Bower et al. 2006; Font et al. 2008, 2011; Lagos et al. 2011) instead uses Dhaloes as its starting point, it is interesting to contrast the properties of haloes defined by these two algorithms.

As described in Section 2, FoF haloes are decomposed by SUBFIND into subhaloes and those are then regrouped into Dhaloes. Hence for every FoF halo, we can find its matching Dhalo by finding which Dhalo contains the most massive subhalo from the FoF group. We can perform this matching the other way round by finding the FoF halo containing the most massive subhalo from the Dhalo. In the cases where the most massive subhalo of an FoF halo is also the most massive subhalo of a Dhalo, these two matching procedures produce identical associations. We refer to such cases as bijective matches.

Before comparing the properties of this subset of bijectively matched Dhaloes and FoF haloes, we first quantify how representative they are by looking at the fraction of each set of haloes that have these bijective matches. The two upper curves in Fig. 1 show the dependence of the bijective fraction of Dhaloes on Dhalo mass and FoF haloes on FoF mass. The first thing to note is that the fraction of bijectively matched Dhaloes is large, being 90 per cent or greater over the full range from 10^8 to $10^{14}\ h^{-1}\ {\rm M}_\odot$ and so to a first approximation there is a good correspondence between FoF and Dhaloes. Above $3 \times 10^{10}\ h^{-1}\ {\rm M}_\odot$, about 10 per cent of the Dhaloes do not have a bijective match which means they instead represent secondary fragments of more massive FoF haloes that the Dhalo algorithm has split into two or more subhaloes. Below $3 \times 10^{10}\ h^{-1}\ {\rm M}_\odot$, this non-bijective fraction drops indicating that lower mass FoF haloes are less likely to be split into two or more comparable mass Dhaloes. This behaviour is consistent with the results of Lukic et al. (2009), who found that 15–20 per cent of FoF haloes are irregular structures that have two or more major components linked together by low-density bridges and that this fraction is an increasing function of halo mass. This is also to be expected in the hierarchical merging picture as the most massive haloes formed most recently and so are the least dynamically relaxed.

[3] As a subhalo can, by definition, only belong to one Dhalo and as particles can only belong to one SUBFIND subhalo, this means that Dhaloes are exclusive in the sense that no particles can belong to more than one Dhalo.

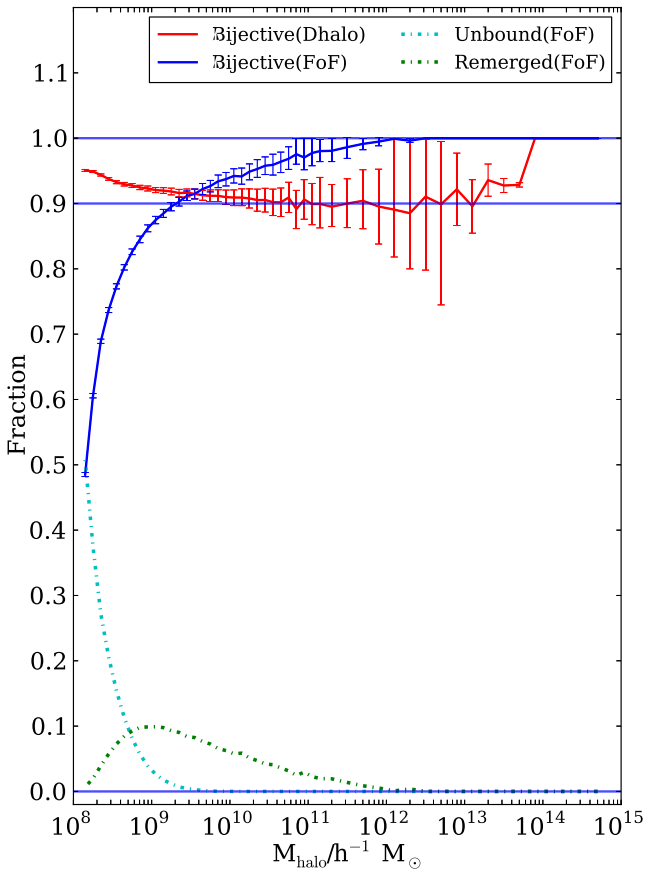

Figure 1. The upper two curves, with bootstrap error bars, show the fraction of Dhalo (red) and FoF haloes (blue) in the MSII catalogues that have a bijective (a unique one-to-one) match as a function of their respective Dhalo or FoF halo mass. The lower two curves show the fraction of FoF haloes that do not contain a self-bound substructure (cyan) and the fraction whose main subhaloes are remerged by the Dhalo algorithm to form part of a more massive Dhalo (green).

For the FoF haloes with mass above $10^{12}\ h^{-1}\ \mathrm{M}_\odot$, the bijectively matched fraction is unity, indicating that the most massive subhalo of such FoF haloes together with the subhaloes embedded within it always gives rise to a Dhalo. Below $10^{12}\ h^{-1}\ \mathrm{M}_\odot$, the bijective fraction begins to decrease steadily with decreasing mass. This happens because as the FoF mass decreases, there is an increasing probability that the progenitor of this FoF halo has previously passed through a more massive neighbouring halo and this results in the Dhalo algorithm remerging the FoF halo with its more massive neighbour. This fraction of FoF haloes that are remerged to form part of a more massive Dhalo is shown by the green curve in Fig. 1. As one approaches $10^{8}\ h^{-1}\ \mathrm{M}_\odot$ ($\sim$15 particles), the bijective fraction plummets as at very low masses many of the FoF haloes are not self-bound and so do not contain any subhaloes from which to build a Dhalo. The fraction of FoF haloes which do not contain a self-bound substructure is shown by the cyan curve in Fig. 1 and can be seen to reach 50 per cent at an FoF mass of 20 particles.

3.1.1 Virial masses

It is conventional to define the virial mass, $M_{\rm vir}$, and associated virial radius, $r_{\rm vir}$, of a dark matter halo using a simple spherical overdensity criterion centred on the potential minimum,

$$M_{\rm vir} = \frac{4}{3}\pi\Delta\,\rho_{\rm crit}\,r_{\rm vir}^3, \quad (1)$$

where $\rho_{\rm crit}$ is the cosmological critical density and Δ is the specified overdensity. In applying this definition, we adopt $\Delta = 200$ and include all the particles inside this spherical volume, not only the particles grouped by the FoF or Dhalo algorithm, to define the enclosed mass, M_{200}, and associated radius r_{200}. This choice is largely a matter of convention but has been shown to roughly correspond to boundary at which the haloes are in approximate dynamical equilibrium (e.g. Cole & Lacey 1996).

If the halo finding algorithm has succeeded in partitioning the dark matter distribution into virialized haloes, we would expect to see a good correspondence between the grouped mass of the halo and M_{200}. For instance, as FoF haloes are essentially bounded by an isodensity contour, whose value is set by the linking parameter (Davis et al. 1985), then if they have relaxed quasi-spherical configuration a tight relation between $M_{\rm halo}$ and M_{200} is inevitable. The only way $M_{\rm halo} \gg M_{200}$ is if the halo has multiple components which have been spuriously linked together as illustrated in the typical example shown in the lower panels of Fig. 4.[4] $M_{\rm halo} \ll M_{200}$ could indicate the cases where the group finder has split a virialized object into small fragments. Hence, it is interesting to look at the distribution of $M_{\rm halo}/M_{200}$ for both the FoF and Dhalo algorithms to simply see how $M_{\rm halo}$ compares to the conventional M_{200} definition of halo mass and to give an indication of the frequency of over linking and fragmentation.

The two panels of Fig. 2 quantify the distribution of $M_{\rm halo}/M_{200}$ for both the standard FoF haloes and for haloes defined by the Dhalo algorithm. We immediately see that the distribution is much tighter for the Dhalo definition than for FoF haloes. For FoF haloes, 5 per cent of the haloes have $M_{\rm FOF}/M_{200} \gtrsim 2$ and 1 per cent have $M_{\rm FOF}/M_{200} \gtrsim 3$. In contrast, for Dhaloes only 5 per cent have $M_{\rm Dhalo}/M_{200} \gtrsim 1.5$ and less than 1 per cent have $M_{\rm Dhalo}/M_{200} > 2$. In the Dhalo panel, only Dhaloes that are bijectively matched with FoF haloes are included. Since such pairs of haloes contain the same most massive subhalo, the centres used for calculating M_{200} are identical and result in the same M_{200}. Furthermore, since Fig. 1 indicates that all FoF haloes more massive than $10^{12}\ h^{-1}\ \mathrm{M}_\odot$ have a bijectively matching Dhalo, then above $10^{12}\ h^{-1}\ \mathrm{M}_\odot$ we are comparing the same population of haloes and using the same values of M_{200}. Consequently, the wider distribution of $M_{\rm halo}/M_{200}$ for FoF is directly caused by the wider spread in $M_{\rm FoF}$ masses. For the cases where $M_{\rm FoF} \gg M_{200}$, there is one or more substantial components of the FoF halo that lies outside r_{200}. We will see in Fig. 4 that these are generally secondary mass concentrations that are linked by tenuous bridges of quite diffuse material. The Dhaloes have a tighter distribution of $M_{\rm halo}/M_{200}$ as in this algorithm these secondary concentrations are successfully split off and result in separate distinct Dhaloes.

Our results for FoF haloes are consistent with earlier investigations. Harker et al. (2006), Evrard et al. (2008) and Lukic et al. (2009) found that approximately 80–85 per cent of FoF haloes are isolated haloes while 15–20 per cent of FoF haloes have irregular

[4] These grossly non-virialized multicomponent systems are not always detected by more often used relaxation criteria (Neto et al. 2007; Power, Knebe & Knollmann 2012, and see Section 4.2), as such criteria focus on the mass within r_{200} which can be in equilibrium even if diffusely linked to secondary mass concentrations.

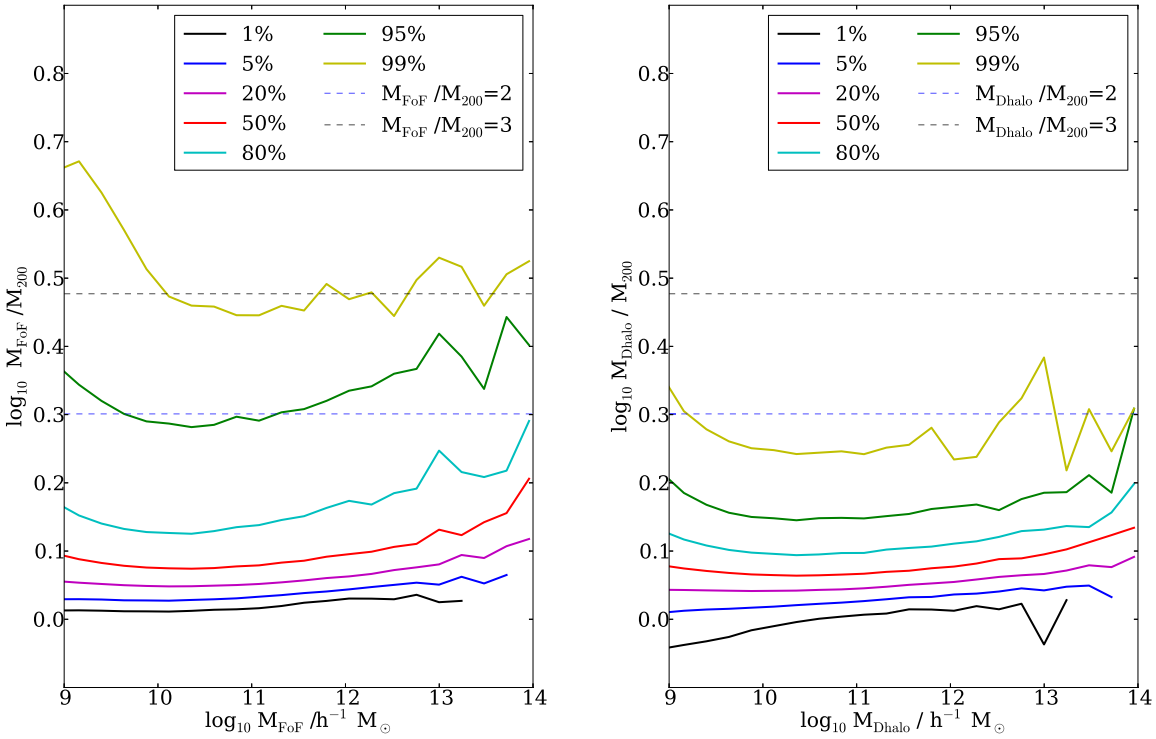

Figure 2. The left-hand panel shows the median, 1, 5, 20, 80, 95 and 99 percentiles of the distribution of the mass ratios between FoF halo mass, $M_{\rm FoF}$, and virial mass, M_{200}, as a function of FoF halo mass for haloes identified using the FoF group finder. The right-hand panel shows the same percentiles for the distribution of the mass ratio between Dhalo mass, $M_{\rm Dhalo}$, and virial mass, M_{200}, as a function of Dhalo mass for haloes identified using the Dhalo group finder. The blue dashed line in both panels shows where $M_{\rm Halo}/M_{200} = 2.0$ and the black one $M_{\rm Halo}/M_{200} = 3.0$.

morphologies, most of which are described in Lukic et al. (2009) as 'bridged haloes'. The distribution of $M_{\rm FoF}/M_{200}$ for 'bridged haloes' given in fig. 7 of Lukic et al. (2009) is very similar to the 20 per cent tail of our distribution above $M_{\rm FoF}/M_{200} = 1.5$, while the isolated haloes in Lukic et al. (2009) have a distribution similar to the remaining 80 per cent of our distribution.

3.1.2 Mass scatter plots

We now turn to directly comparing the mass assigned to FoF haloes and their corresponding Dhaloes. Fig. 3 compares the distributions of these two masses and their ratio for bijectively matched FoF and Dhaloes, i.e. haloes which contain the same most massive subhalo. First, we see that the median of the distribution is very close to the one-to-one line. Furthermore, on one side the distribution cuts off very sharply with far fewer than 1 per cent of haloes having FoF masses significantly lower than their corresponding Dhalo mass. In principle, $M_{\rm Dhalo} > M_{\rm FoF}$ can occur as one aspect of the Dhalo algorithm is that it includes satellite subhaloes that previously passed through the main halo even if they are now sufficiently distant so as not to be linked into the corresponding FoF halo. However, such subhaloes are typically much less massive than the main subhalo and the mass gained in this way is outweighed by other sources of mass loss. On the other side of the distribution, there is a significant tail of haloes for which $M_{\rm FoF} > M_{\rm Dhalo}$. We see that approximately 5 per cent have $M_{\rm FoF} > 1.5M_{\rm Dhalo}$ and 1 per cent have $M_{\rm FoF} > 2M_{\rm Dhalo}$. These fractions are largely independent of the Dhalo mass. The main reason for this tail is the presence of FoF haloes that have a significant secondary mass concentration, often linked by a low-density bridge, that the Dhalo algorithm succeeds in splitting off. For these bijectively matched haloes, $M_{\rm FoF}$ is unlikely to significantly exceed $2M_{\rm Dhalo}$ as if a single secondary mass concentration had a subhalo of mass greater than that of the most massive subhalo in the Dhalo we would not have a bijective match. However, in rare instances $M_{\rm FoF} > 2M_{\rm Dhalo}$ can occur when the FoF halo contains several massive secondary mass concentrations.

To illustrate the relationship between FoF and Dhaloes, we show three examples in Fig. 4 that have been chosen to be representative of different points in the $M_{\rm FoF}$–$M_{\rm Dhalo}$ distribution. The halo shown in the top row is representative of the majority of cases, namely those with $M_{\rm FoF} \approx M_{\rm Dhalo}$. Here the only particles from the FoF halo that are not included in the Dhalo are a diffuse cloud of unbound particles and the particles in a couple of subhaloes whose centres lie outside twice the half-mass radius of the main subhalo. We stress that these small differences are what is typical for corresponding FoF and Dhaloes.

The middle row of Fig. 4 shows an example where $M_{\rm FoF}/M_{\rm Dhalo} = 1.5$, which corresponds to the 95th percentile of the distribution shown in Fig. 3. Here the FoF halo is split into three well-separated Dhaloes. The main Dhalo is dominant, but there two secondary Dhaloes, one a lot more massive than the other, lying outside the r_{200} of the main Dhalo. For the purposes of semi-analytic galaxy formation models such as GALFORM, the three separate haloes given by the Dhalo definition are clearly a better description than the single FoF halo as one would not expect the gas reservoirs associated with these distinct haloes to have merged at this stage and so each should be able to provide cooling gas to their respective central galaxies.

The bottom row of Fig. 4 shows a rare example with $M_{\rm FoF}/M_{\rm Dhalo} \approx 2$, the 99th percentile of the distribution, in which a single FoF halo is split into several substantial Dhaloes. In this and the previous example, the FoF halo is clearly far from spherical and a large

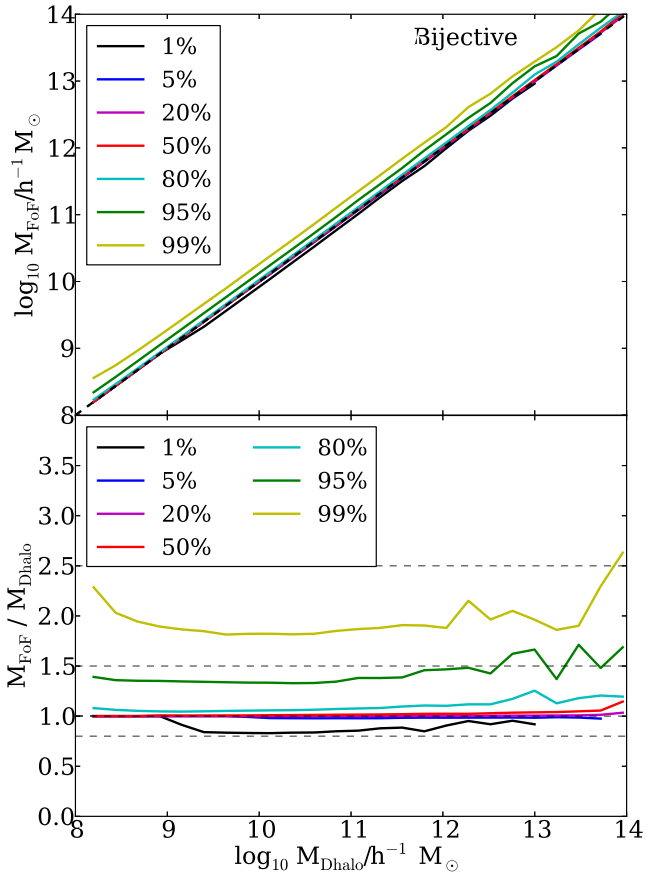

Figure 3. In the top panel, the 1, 5, 20, 50, 80, 95 and 99 percentiles of the distribution of FoF halo mass, $M_{\rm FoF}$, are plotted against $M_{\rm Dhalo}$ for the bijectively matched pairs of haloes. In the bottom panel, the same percentiles of the distribution of the mass ratio $M_{\rm FoF}/M_{\rm Dhalo}$ are plotted as a function of Dhalo mass. The black dashed lines are where $M_{\rm FoF}/M_{\rm Dhalo} = 0.8$, 1, 1.5 and 2.5.

proportion of the FoF halo mass lies outside the virial radius that is defined by centring on the potential minimum of the most massive substructure. Clearly, characterizing such haloes by an NFW profile fit just to the mass within the virial radius would be an inadequate description of the halo. In fact, in most studies of halo concentrations, including our analysis present in Section 4.2, these haloes would be deemed to be unrelaxed and excluded from subsequent analysis. In contrast, the Dhaloes in each of the examples presented are much closer to being spherical with only a small amount of mass outside their respective virial radii. Each of the primary Dhaloes in Fig. 4, including the one in the bottom panel, is sufficiently symmetrical and virialized to pass the relaxation criteria that we employ in Section 4.2 even though the corresponding FoF haloes in the bottom two panels are not.

In the example shown in the bottom row of Fig. 4, we also see the case of a Dhalo that has two distinct components. Here the two clumps of black points are a single Dhalo due to the fact that they passed directly through each other at a redshift $z = 0.89$. This extreme example must have been a high-speed encounter and so any galaxies they contained would have been unlikely to merge, but their extended hot gas distributions would have interacted and possibly merged. It is for this reason that it is useful in the semi-analytic models to associate them as a single halo.

The Dhalo algorithm quite frequently merges several FoF haloes together into a single Dhalo as a consequence of the way it avoids splitting up subhaloes which at an earlier timestep were in a single Dhalo. However, unlike the extreme example we have just seen that the typical masses of subhaloes which pass through a Dhalo and then emerge to once again become a distinct FoF halo are much lower than the mass of the main FoF halo. This is illustrated in Fig. 5, where we show the particles of three typical Dhaloes of a range of masses colour coded by their FoF halo membership. In each case, we immediately see that the vast majority of the Dhalo particles also belong to the (bijectively) matched FoF halo. However, in addition there are isolated clumps of particles in the outskirts of each Dhalo which belong to much smaller distinct FoF haloes. There are also similar nearby clumps of particles which due to surrounding diffuse material are linked into the main FoF halo. In all cases, each of these clumps is typically less than 1 per cent of the mass of the main halo. From the perspective of semi-analytic galaxy formation models, it makes sense to treat each of these clumps equally. For instance, they have all been within twice the half-mass radius of the main Dhalo and could therefore have been ram pressure stripped of their diffuse gaseous haloes. In GALFORM, satellite galaxies move with the subhalo within which they formed (or if the descendant of the subhalo drops below the 20 particle threshold with the particle that was previously the potential minimum of its subhalo) and so the satellite galaxy positions reflect the spatial distribution of these subhaloes even if they move far from the halo with which they are associated.

3.2 Non-bijective FoF and Dhalo matches

So far we have just compared FoF–Dhalo pairs which form a bijective match, that is their most massive subhaloes are identical. However, there other cases such as the examples of secondary Dhaloes in Fig. 4 in which the main subhalo of the Dhalo is not the most massive subhalo in the corresponding FoF halo and conversely examples such as the secondary FoF haloes in Fig. 5 in which the main subhalo of the FoF halo is not the most massive subhalo in the corresponding Dhalo. We will refer to the former set of matches as Dhalo in FoF halo and the latter as FoF in Dhalo matches. Note that the bijective matches are a subset of both of these sets, i.e. they are the intersection of the two sets of matches. To have a complete census of the correspondence between FoF and Dhaloes, it is important that we include non-bijectively matched haloes in our comparison. We compare the Dhalo to FoF halo masses for these two sets of pairings in Figs 6 and 7.

The left-hand panels of Fig. 6 show for all Dhalo in FoF halo matches the dependence of the mass, $M_{\rm Dhalo}$, and the mass ratio, $M_{\rm Dhalo}/M_{\rm FoF}$, on the FoF halo mass. The right-hand panel shows the same quantities but only for secondary Dhalo in FoF halo haloes, i.e. excluding the bijective matches. Focusing first on the right-hand panels, we see that the percentiles of the distribution of secondary $M_{\rm Dhalo}$ values are all horizontal lines at high $M_{\rm FoF}$, indicating that in this regime the distribution of $M_{\rm Dhalo}$ is independent of $M_{\rm FoF}$. This suggests that the secondary Dhaloes that are linked into high-mass FoF haloes by bridges of diffuse material are essentially drawn at random from the Dhalo population. We note that in this way the FoF halo can be hundreds or more times more massive than many of the Dhaloes it contains. In these same panels, we see that at lower masses the distribution of Dhalo masses is sharply truncated at $M_{\rm Dhalo} = M_{\rm FoF}/2$. This is essentially by construction as if a Dhalo with mass greater than $M_{\rm FoF}/2$ were linked into the FoF halo, then its most massive subhalo would very likely to be the most massive subhalo of the whole FoF halo and hence there would be a bijective match and this pairing would be excluded from this plot.

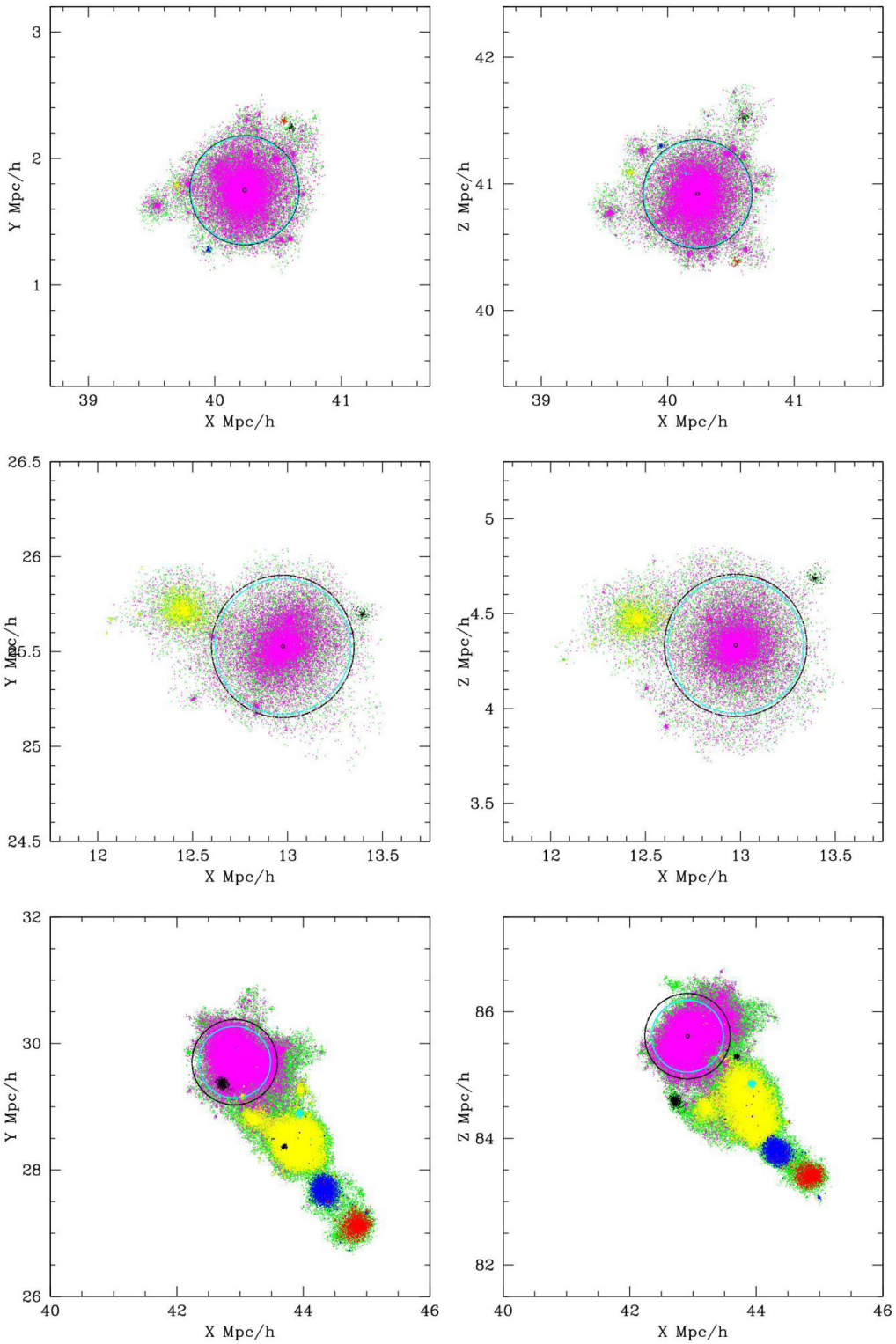

Figure 4. Three examples of the relationship between FoF haloes and Dhaloes. In each panel, all the points plotted are from a single FoF halo. First, all the FoF particles were plotted in green and then subsets belonging to specific Dhaloes were overplotted. The magenta points are those belonging to the bijectively matched Dhaloes. Other colours are used to indicate particles belonging to other non-bijective Dhaloes with a unique colour used for each separate Dhalo. Two projections of each halo are shown. The left-hand panels show the X–Y and right the X–Z plane. The black circle marks r_{200} of the FoF halo and the cyan circle marks twice the half-mass radius of the main subhalo of the FoF halo. The top row shows a typical case where $M_{\rm FoF} \approx M_{\rm Dhalo}$. Here $M_{\rm FoF} = 2.6 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$, $M_{200} = 1.9 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$ and $r_{200} = 0.43\,h^{-1}$ Mpc. The middle panel shows an example where the mass ratio $M_{\rm FoF}/M_{\rm Dhalo} = 1.5$ with $M_{\rm FoF} = 1.7 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$, $M_{200} = 1.2 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$ and $r_{200} = 0.375\,h^{-1}$ Mpc. The bottom row shows an extreme example where $M_{\rm FoF} \gg M_{\rm Dhalo}$ and the FoF halo is split into many Dhaloes. Here $M_{\rm FoF} = 1.4 \times 10^{14}\,h^{-1}\,{\rm M}_{\odot}$, $M_{200} = 7.1 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$ and $r_{200} = 0.67\,h^{-1}$ Mpc.

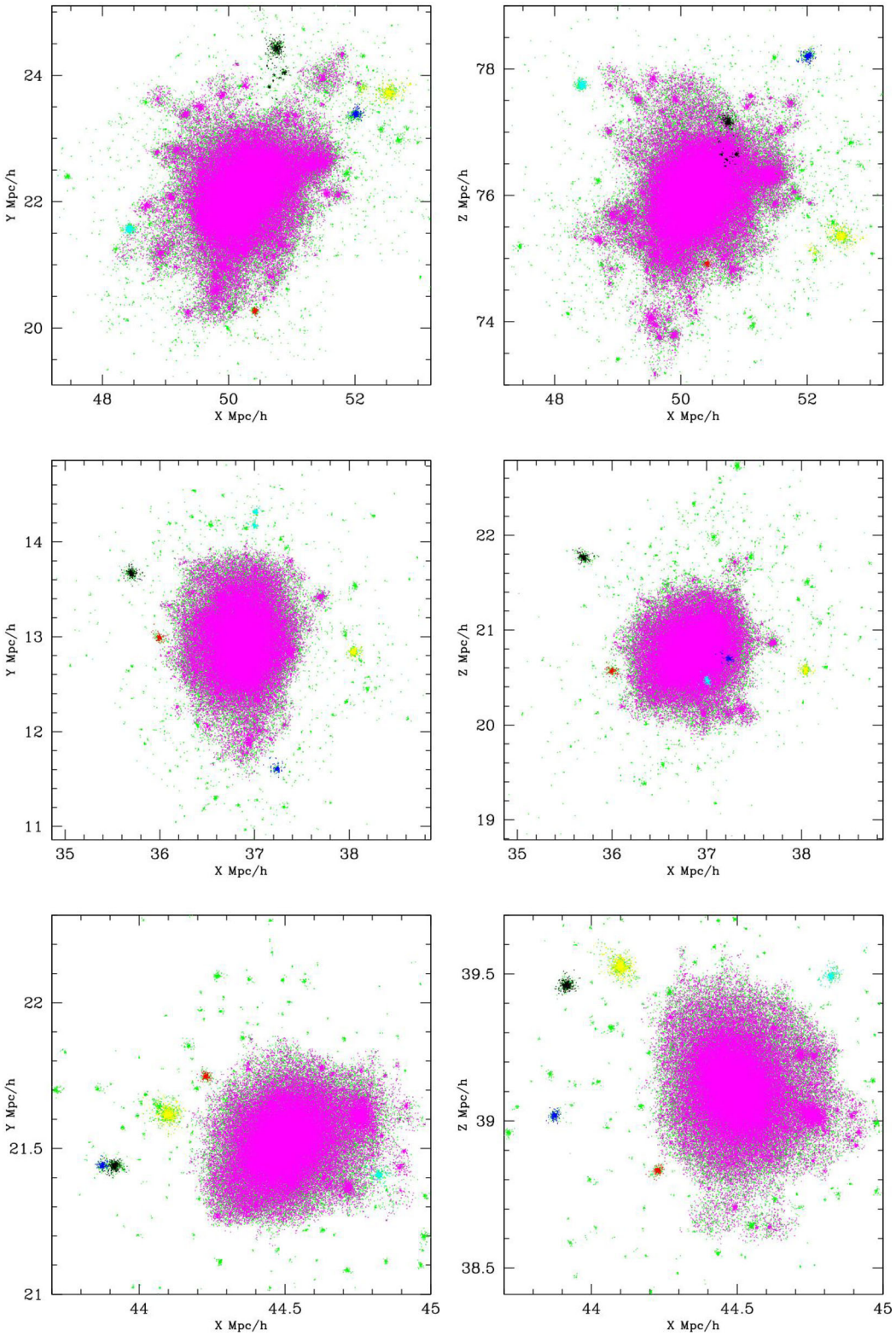

Figure 5. Examples of three typical Dhaloes showing how a single Dhalo can be composed of more than one FoF halo. In each panel, all the points plotted are from a single Dhalo. First, all the Dhalo particles were plotted in green and then subsets belonging to specific FoF haloes were overplotted. The magenta points are those belonging to the bijectively matched FoF halo. Other colours are used to indicate particles belonging to other FoF haloes with a unique colour used for each separate FoF halo. Two projections of each halo are shown. The left-hand panels show the X–Y and right the X–Z plane. From top to bottom, the Dhalo masses of these examples are $M_{\rm Dhalo} = 4.2 \times 10^{14}\,h^{-1}\,{\rm M}_{\odot}$, $M_{\rm Dhalo} = 6.8 \times 10^{13}\,h^{-1}\,{\rm M}_{\odot}$ and $M_{\rm Dhalo} = 5.4 \times 10^{12}\,h^{-1}\,{\rm M}_{\odot}$. In all cases, the majority of the Dhalo mass is contained in the single bijectively matched FoF halo, and the secondary FoF haloes are typically 100 times less massive.

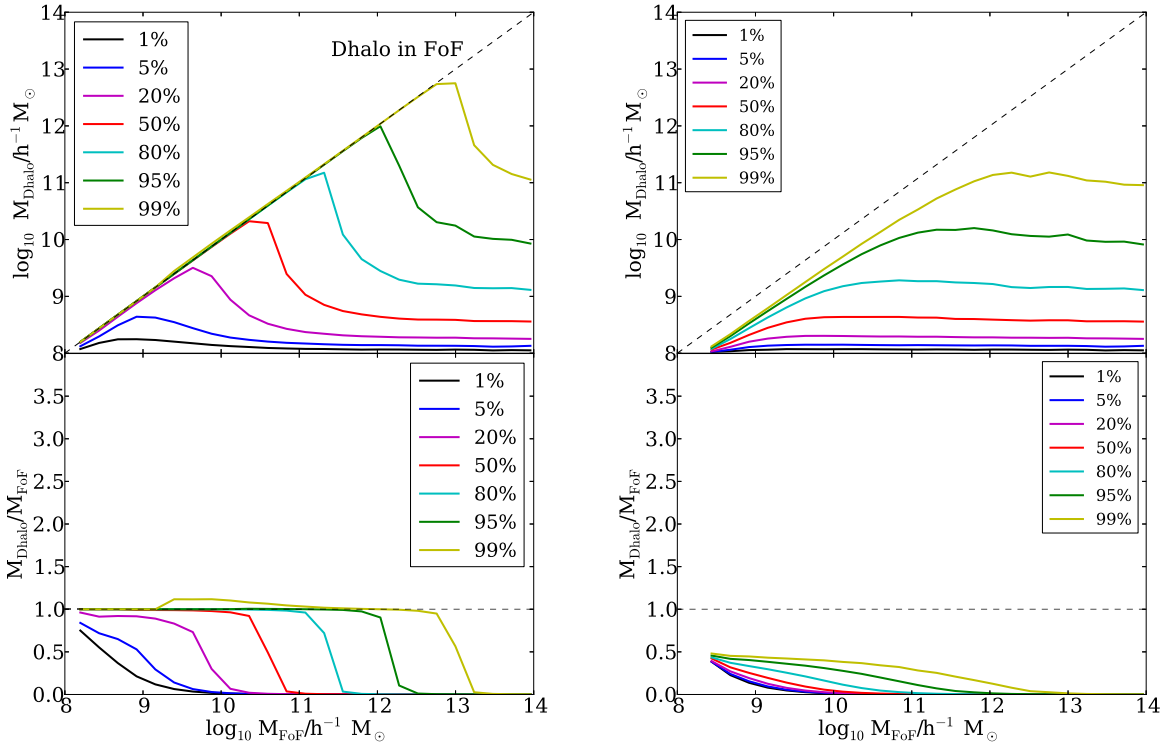

Figure 6. In the left-hand panels, we plot the median, 1, 5, 20, 80, 95 and 99 percentiles of the distribution of Dhalo mass, $M_{\rm Dhalo}$ (upper), and mass ratio $M_{\rm Dhalo}/M_{\rm FoF}$ (lower) against $M_{\rm FoF}$ for all the Dhalo matches to each FoF halo. The black dashed lines in each panel mark where $M_{\rm Dhalo}/M_{\rm FoF} = 1$. In the right-hand panel, we plot the same quantities but only for secondary Dhaloes in each FoF halo.

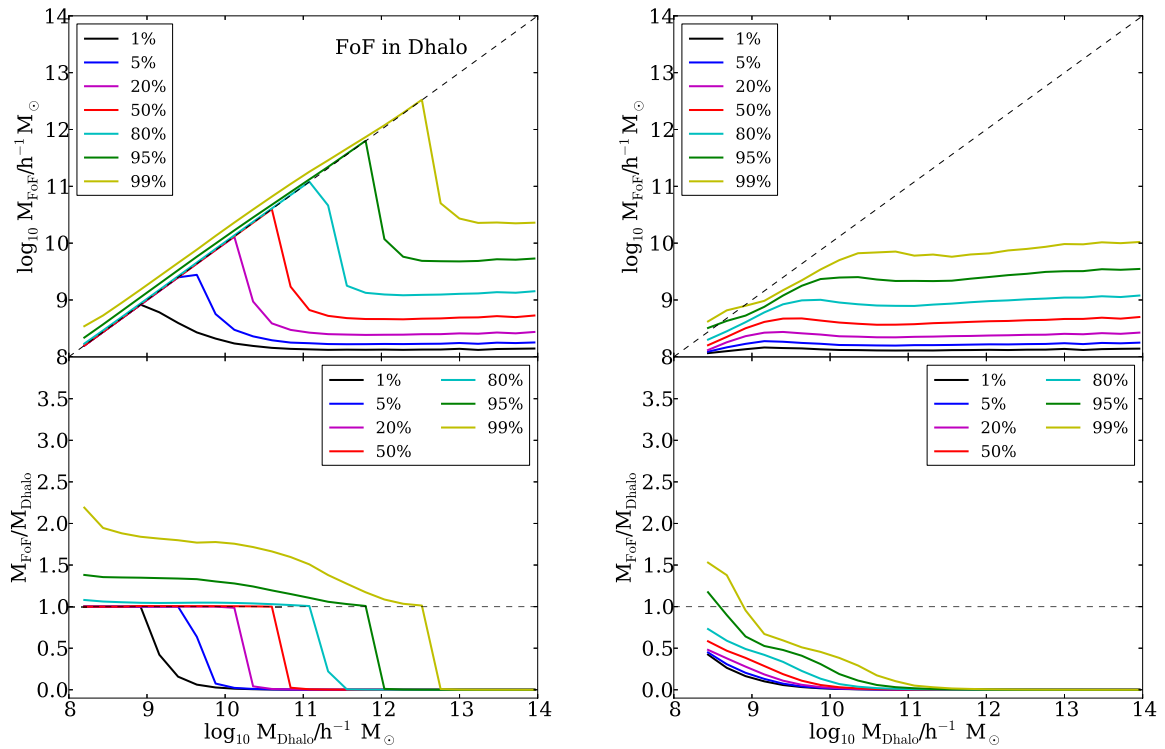

Figure 7. As Fig. 6 but with the role of FoF and Dhalo reversed. In the left-hand panels, we plot the median, 1, 5, 20, 80, 95 and 99 percentiles of the distribution of FoF halo mass, $M_{\rm FoF}$ (upper), and mass ratio $M_{\rm FoF}/M_{\rm Dhalo}$ (lower) against $M_{\rm Dhalo}$ for all the FoF halo matches to each Dhalo. The black dashed lines in each panel mark where $M_{\rm FoF}/M_{\rm Dhalo} = 1$. In the right-hand panel, we plot the same quantities but only for secondary FoF in each Dhalo.

The left-hand panels of Fig. 6, which includes the bijective matches, show a more complex distribution. However, it can be easily understood as resulting from the superposition of the distribution from the right-hand panel with the distribution of bijective matches shown in Fig. 3. At very low masses, most FoF haloes contain only a single resolved subhalo and so the FoF halo cannot be split into multiple Dhaloes and so the overall distribution is dominated by the bijective matches resulting in a tight correlation between $M_{\rm Dhalo}$ and $M_{\rm FoF}$. With increasing FoF mass, there are more and more secondary Dhaloes per FoF halo. They increasingly dominate over the bijective matches and so the contours tend to their values in the right-hand panel.

Fig. 7 shows the distribution of FoF halo mass for the FoF in Dhalo matches. Again the right-hand panels show the distribution for just the secondary matches while the left-hand panels also include the primary or bijective matches. Comparing the right-hand panels of Figs 6 and 7, we see that the corresponding contours are shifted to lower masses. Thus, it is rarer for a Dhalo to contain a massive secondary FoF halo than it is for an FoF halo to contain a massive secondary Dhalo. The secondary Dhaloes arise from the remerging step in the Dhalo algorithm whereby two subhaloes that have passed through each other (the smaller has come within twice the half-mass radius of the larger) are deemed thereafter always to be part (or satellite components) of the same Dhalo even if they subsequently separate sufficiently to become distinct FoF haloes. This occurs reasonably frequently, but as in the examples shown in Fig. 4 the secondary FoF haloes are typically much less massive than the primary and contribute little to the total mass of the halo. Interestingly, the near-horizontal contours in the upper-right hand panel Fig. 7 indicate that the mass distribution of this population of secondary FoF haloes is approximately independent of $M_{\rm Dhalo}$ for high Dhalo masses. As these FoF haloes are often heavily stripped by their passage through the main Dhalo, this is not a trivial result. The contours begin to dip at lower masses reflecting the fact it is unlikely for a matched FoF halo to have a mass greater than about one half of $M_{\rm Dhalo}$ without it being the primary or bijective match. This expectation is violated for $M_{\rm Dhalo} < 10^9\,h^{-1}\,{\rm M}_\odot$, but this is a resolution effect because at such low masses secondaries with $M_{\rm FOF} \ll M_{\rm Dhalo}$ fall below the 20 particle limit of the catalogue and so their absence biases the distribution towards higher ratios.

The left-hand panels of Fig. 7 are for all the matches of FoF in Dhalo, including the bijective matches. These distributions can be understood as a superposition of the distributions in the right-hand panels with the distribution for bijective matches shown in Fig. 3. At low masses, the bijective halo matches dominate whereas at large $M_{\rm Dhalo}$ there are many FoF haloes matched to each Dhalo. Thus, for example, at $M_{\rm Dhalo} \approx 10^{10.5}\,h^{-1}\,{\rm M}_\odot$, we transition from 50 per cent of the matched FoF haloes being primary to 50 per cent of them being much lower mass ($M_{\rm FoF} \approx 10^{8.7}\,h^{-1}\,{\rm M}_\odot$) secondary FoF haloes.

In Section 3.1.1, we examined the distribution of the $M_{\rm Dhalo}/M_{200}$ ratio for the bijectively matched haloes. We are also interested in this distribution for the non-bijective Dhaloes shown in Fig. 8. We immediately notice that the distribution is shifted towards lower values than the corresponding distribution for the bijective haloes shown in Fig. 2. The origin of this shift can be understood by reference to Fig. 9 which shows an example of an FoF halo which is split into several Dhaloes. The Dhalo whose particles are plotted in magenta is the bijective match of the FoF halo and the Dhaloes plotted in other colours are non-bijective matches. The black circles in Fig. 9 show the location of r_{200} for each of the Dhaloes, while the other circles show the location of the half-mass radius of each

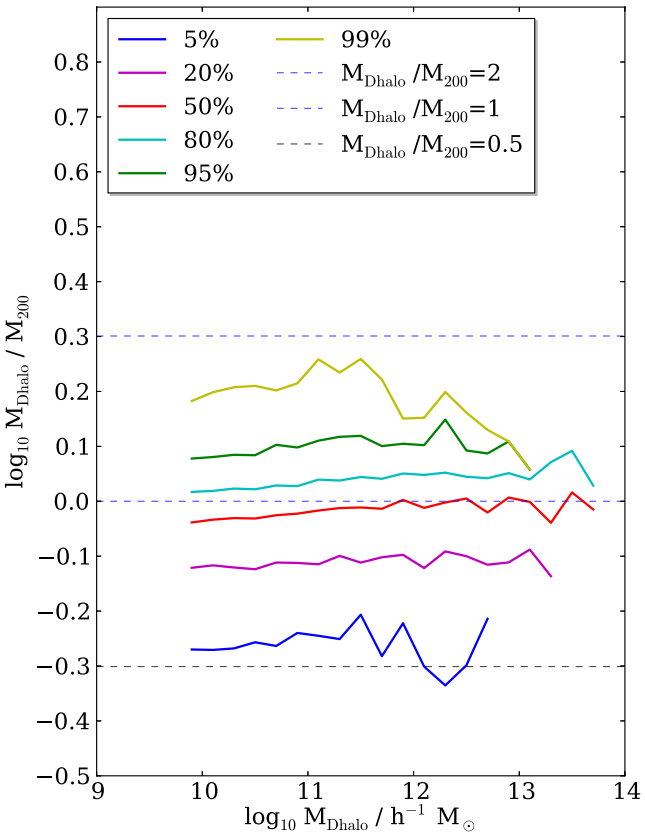

Figure 8. Like the right-hand panel of Fig. 2, but for non-bijective Dhaloes. The curves show the median, 5, 20, 80, 95 and 99 percentiles of the ratio between the Dhalo mass, $M_{\rm Dhalo}$, and the virial mass, M_{200}. The horizontal dashed lines indicate $M_{\rm Dhalo}/M_{200} = 0.5$, 1.0 and 2.0.

Dhalo. For bijectively matched Dhaloes, the majority of which are isolated, r_{200} is typically slightly smaller than the half-mass radius. In contrast, we see in Fig. 9 that for many of the non-bijectively matched Dhaloes, the half-mass radius is much smaller than r_{200}. This is a consequence of the SUBFIND algorithm which determines the extent of a subhalo by finding saddle points in the density distribution (Springel et al. 2001). Hence, as a subhalo enters a dense environment, the mass assigned to it by SUBFIND is decreased. This environmentally dependent effect both lowers $M_{\rm Dhalo}$ relative to M_{200} and increases the scatter in this relation.

4 STATISTICAL PROPERTIES OF DHALOES

Having thoroughly compared individual Dhaloes with their corresponding FoF haloes, we now turn to the statistical properties of the Dhaloes. We first look at the Dhalo mass function and then the statistics of their density profiles as characterized by fitting NFW profiles (Navarro et al. 1995, 1996; Navarro, Frenk & White 1997).

4.1 The Dhalo mass function

For many applications, it is extremely useful to have an analytic description of the number density of haloes as a function of the halo mass. A relevant example for us is when semi-analytic galaxy formation models are constructed using Monte Carlo methods (Cole et al. 2000; Parkinson, Cole & Helly 2008) of generating dark matter merger trees. In this case, in order to construct predictions of galaxy luminosity functions or any other volume-averaged quantity

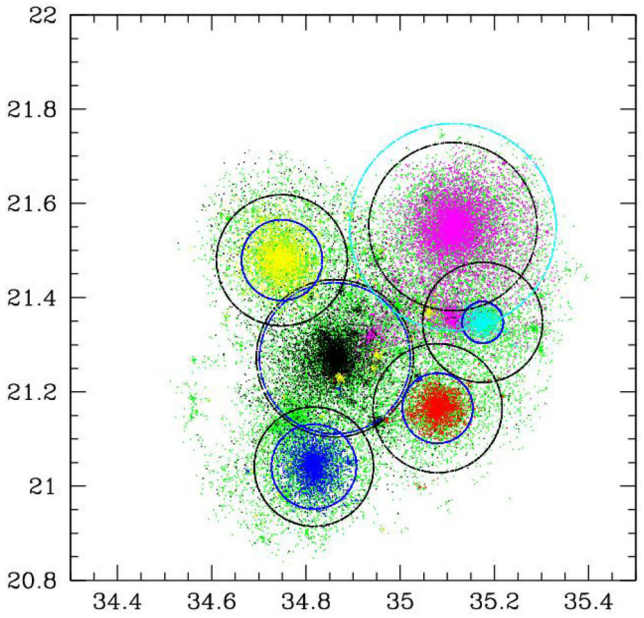

Figure 9. An example of one FoF halo split by the Dhalo algorithm into several Dhaloes. All the points plotted are from a single FoF halo. First, all the FoF particles are plotted in green and then subsets belonging to specific Dhaloes are overplotted. The magenta points are those belonging to the bijectively matched Dhalo. Other colours are used to indicate particles belonging to other Dhaloes with a unique colour used for each separate Dhalo. The black circle around the magenta points marks r_{200} of the FoF halo and is also the r_{200} of the bijective Dhalo. The concentric cyan circle marks twice the half-mass radius of this main subhalo. The other black circles show r_{200} locations for the non-bijective Dhaloes, while the concentric blue circles indicate twice the half-mass radius of the corresponding subhalo.

(Cole et al. 2000; Berlind et al. 2003; Baugh et al. 2005; Bundy, Ellis & Conselice 2005; van den Bosch, Tormen & Giocoli 2005; Giocoli, Pieri & Tormen 2008; Moreno, Giocoli & Sheth 2008; Neistein & Dekel 2008), one needs knowledge of the halo mass function in order to know how many of each type of tree one has per unit volume. It has become common practice to assume that the halo mass function is given by analytic fitting functions which have been fitted to the abundance of haloes found by the FoF or other group finding algorithms (Davis et al. 1985; Lacey & Cole 1994; Knollmann & Knebe 2009) in suites of cosmological *N*-body simulations. Murray, Power & Robotham (2013a,b) compare all the currently proposed fitting functions. In our semi-analytic modelling, we would like to achieve consistent results when using Monte Carlo merger trees or when using merger trees extracted directly from *N*-body simulations using the Dhalo algorithm. Hence, it is important to directly determine the Dhalo mass function and to compare it to such fitting formulae.

We do this in Fig. 10 which compares the Dhalo and FoF mass functions that we measure in the MSII simulations with various analytic prescriptions (Jenkins et al. 2001; Sheth & Tormen 2002; Warren et al. 2006; Reed et al. 2007; Tinker et al. 2008; Watson et al. 2013). The left-hand panel shows the number density of haloes per unit logarithmic interval of mass from the nominal 20 particle mass resolution of the simulation up to $10^{14}\,h^{-1}\,\mathrm{M}_{\odot}$ which is the mass of the most massive haloes in the simulation. In constructing these mass functions, the halo mass we use is simply the aggregated mass of all the particles assigned to each halo. Thus, in the FoF case this is all particles linked to the halo by the FoF algorithm while in the Dhalo case it is the sum of the masses of the subhaloes that compose an individual Dhalo. Also shown in this panel are the predictions of various analytic prescriptions. To evaluate these, we use $\sigma^2(M)$, the variance of the density fluctuations as a function of mass (using a top-hat filter), corresponding to the input power spectrum of the MSII propagated to the output time of the simulation using linear theory. They are all clearly very similar and so in the left-hand panel we expand the dynamic range of the comparison by plotting each mass function divided by the prediction of the Sheth & Tormen (2002) model.

The first thing that we note is that despite the sometimes quite large differences (see Section 3) in the masses of individual FoF and Dhaloes, their two mass functions agree to within 5 per cent for all masses greater than $10^{10}\,h^{-1}\,\mathrm{M}_{\odot}$. In the range $10^{10} \lesssim M_{\rm halo} \lesssim 10^{12.5}\,h^{-1}\,\mathrm{M}_{\odot}$, the Dhalo abundance is approximately 5 per cent higher than FoF haloes as roughly 5 per cent of Dhaloes are secondary members of FoF haloes. In other words, the FoF halo abundance has been suppressed relative to the Dhalo abundance by a fraction of them being composed of two or more Dhaloes that have been linked into one more massive FoF halo by diffuse material or bridges. There is also a competing effect, FoF haloes being remerged into single Dhaloes, which suppressed the Dhalo abundance, but this is a much smaller effect.

Below $10^{10}\,h^{-1}\,\mathrm{M}_{\odot}$, the abundance of FoF halo rises systematically above that of Dhaloes. Between 10^{10} and $8 \times 10^{8}\,h^{-1}\,\mathrm{M}_{\odot}$, this excess increases to about 10 per cent and is caused by FoF haloes that are remerged to become secondary components of larger Dhaloes (see Fig. 1). At lower masses ($\lesssim$100 particles), the sharp upturn in the FoF mass function relative to that of Dhaloes is due to an increasing fraction of the FoF haloes not containing a self-bound subhalo and so having no corresponding Dhalo (see Fig. 1). Thus, this portion of the mass function is strongly affected by the resolution of the simulation.

The Jenkins et al. (2001) fitting formula is within 10 per cent of both the FoF and Dhalo mass functions for masses above $2 \times 10^{10}\,h^{-1}\,\mathrm{M}_{\odot}$. However, below this mass it strongly underpredicts the number density of low-mass haloes. Note that we only plot this fit and that of Watson et al. (2013) over the mass ranges used to constrain them in the original papers. The Watson et al. (2013) mass function is only defined at very high masses where we have poor statistics. It lies somewhat below but is still compatible with our noisy estimates. The Warren et al. (2006) model has the best agreement with our FoF mass function, fitting it well all the way down to 40 particles, beyond which we expect our limited resolution means that our FoF mass function is contaminated by spurious unbound chance groupings of particles. However, the Reed et al. (2007) mass function does a better job of matching the low-mass end of our Dhalo mass function. The Sheth & Tormen mass function is intermediate at low masses between that of Warren et al. (2006) and Reed et al. (2007), but systematically below the other models and our FoF and Dhalo mass functions at high masses, though still only at the 15 per cent level. The Tinker et al. (2008) mass function predicts halo abundances that are about 5–10 per cent higher than Warren et al. (2006) and our estimated FoF abundances.

In summary, the Dhalo and FoF mass functions are very similar and only differ by more than 5 per cent below $10^{10}\,h^{-1}\,\mathrm{M}_{\odot}$. As a result, the established analytic mass function models fit the Dhalo mass function almost as well as they do the standard FoF mass function. The differences between the different analytic fitting formulae are greater than the difference between the FoF and Dhalo mass functions. The Reed et al. (2007) model is a slightly better description of the Dhalo mass function due to it predicting a slightly lower abundance at low masses.

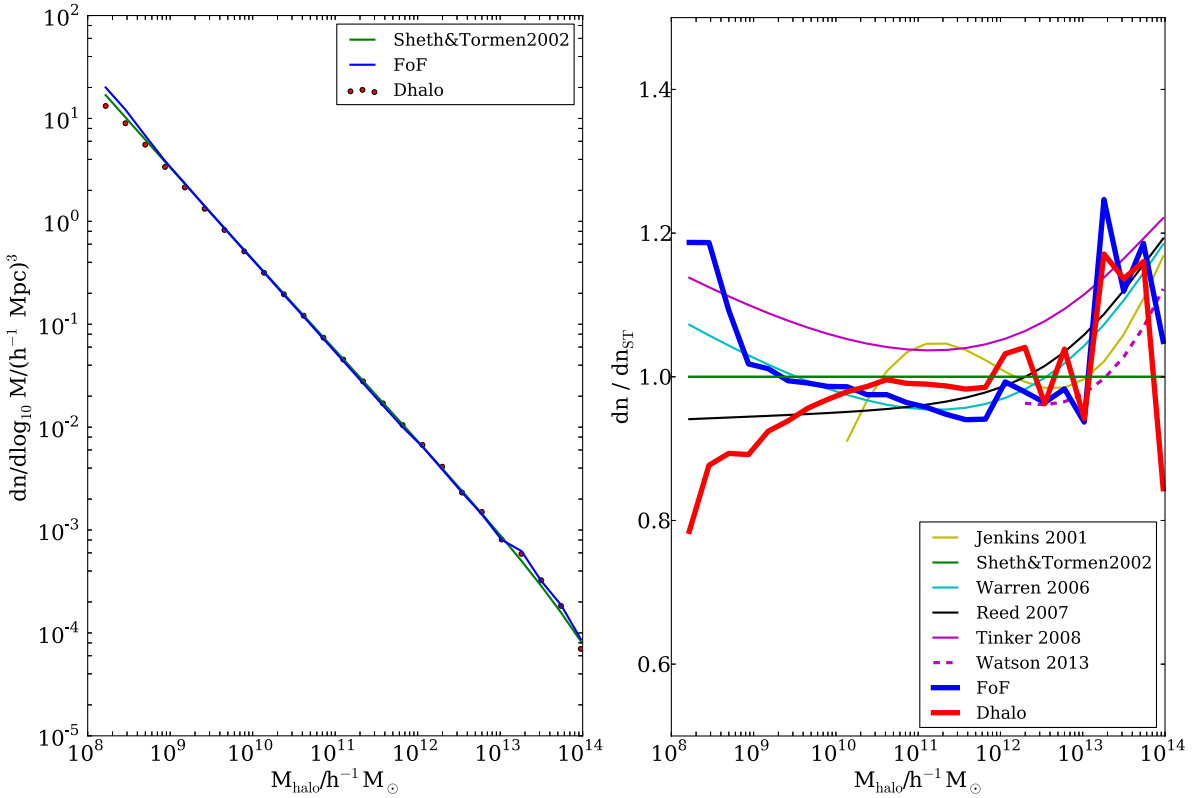

Figure 10. The left-hand panel shows the differential mass functions for both FoF (linking length $b = 0.2$) haloes (blue line) and Dhaloes (red points) in the MSII simulation. We plot this down to $\sim 10^8\, h^{-1}\ \mathrm{M}_\odot$, the mass corresponding to 20 particles in the MSII simulation and we also plot the Sheth & Tormen (2002) mass function as a comparison. To expand the dynamic range, the right-hand panel shows the corresponding prediction of various analytic mass functions (Jenkins et al. 2001; Warren et al. 2006; Reed et al. 2007, Tinker et al. 2008, Watson et al. 2013) as indicated in the legend but now relative to the Sheth & Tormen (2002) prediction. The FoF Dhalo data are now shown as the heavy blue and red lines.

4.2 Density profile fits

We now turn to the density profiles of the haloes as these are an important ingredient in semi-analytic models such as GALFORM where they influence the rate at which gas cools and set the gravitational potential well in which galaxies form. We choose to fit the halo density profiles using NFW (Navarro et al. 1996, 1997) profiles

$$\frac{\rho_{\rm NFW}(r)}{\rho_{\rm crit}} = \frac{\delta_c}{r/r_{\rm s}(1+r/r_{\rm s})^2} \quad (r \le r_{200}), \tag{2}$$

where δ_c is the characteristic density contrast and $r_{\rm s}$ is the scale radius. We define the virial radius, r_{200}, as the radius at which the mean interior density equals 200 times the critical density, $\rho_{\rm crit} = 3H_0^2/(8\pi G)$. The concentration is defined as $c \equiv r_{200}/r_{\rm s}$. The definition of r_{200} implies that δ_c and c must satisfy

$$\delta_c = \frac{200}{3}\frac{c^3}{\ln(1+c) - c/(c+1)}. \tag{3}$$

Our choice of NFW profiles is motivated by their accuracy as a model of CDM haloes (Navarro et al. 1996, 1997), their widespread use and so that our results can be compared to those in Neto et al. (2007), who studied the statistics of NFW concentrations for FoF haloes identified in the Millennium Simulation (Springel 2005). To allow us to compare directly with Neto et al. (2007), we have followed their fitting procedure.

For each halo, we have computed a spherically averaged density profile by binning the halo mass into 32 equally spaced bins in $\log_{10}(r)$ between the virial radius and $\log_{10}(r/r_{200}) = -2.5$, centred on the potential minimum. We fit the two free parameters, δ_c and $r_{\rm s}$, by minimizing the mean square deviation

$$\sigma^2_{\rm fit} = \frac{1}{N_{\rm bin}-1}\sum_i^{N_{\rm bin}} [\log_{10}\rho(r_i) - \log_{10}\rho_{\rm NFW}(r_i|\delta_c, r_{\rm s})]^2 \tag{4}$$

between the binned $\rho(r)$ and the NFW profile. As in Neto et al. (2007), we perform the fit over the radial range $0.05 < r/r_{200} < 1$. In order to be consistent with the original NFW work, we express the results in terms of fitted virial mass, M_{200}, and a concentration, $c_{200} \equiv r_{200}/r_{\rm s}$. We note that while the fitted value of M_{200} used here and the directly measured M_{200} used earlier (e.g. in Fig. 2) are not identical, they in general agree very accurately with an rms scatter of less than 3 per cent.

Neto et al. (2007) distinguished relaxed haloes from haloes that were not in dynamical equilibrium due to recent or ongoing mergers. They found that relaxed haloes were well fitted by NFW profiles while the profiles of unrelaxed haloes were lumpier and yielded poorer fits with systematically lower concentrations. Hence, to compare to Neto et al. (2007), we use the following three objective criteria to assess whether a halo has reached equilibrium (Neto et al. 2007; Gao et al. 2008; Power et al. 2012).

(i) The fraction of mass in resolved substructures whose centres lie inside r_{200}: $f_{\rm sub} = \sum_{i\neq 0}^{N_{\rm sub}} M_{{\rm sub},i}/M_{200}$. We require $f_{\rm sub} < 0.1$ for relaxed haloes.

(ii) The centre of mass displacement, i.e. the difference between the position of the potential minimum and the centre of mass, $s = |r_{\rm c} - r_{\rm cm}|/r_{200}$ (Thomas et al. 2001). Note that the centre of

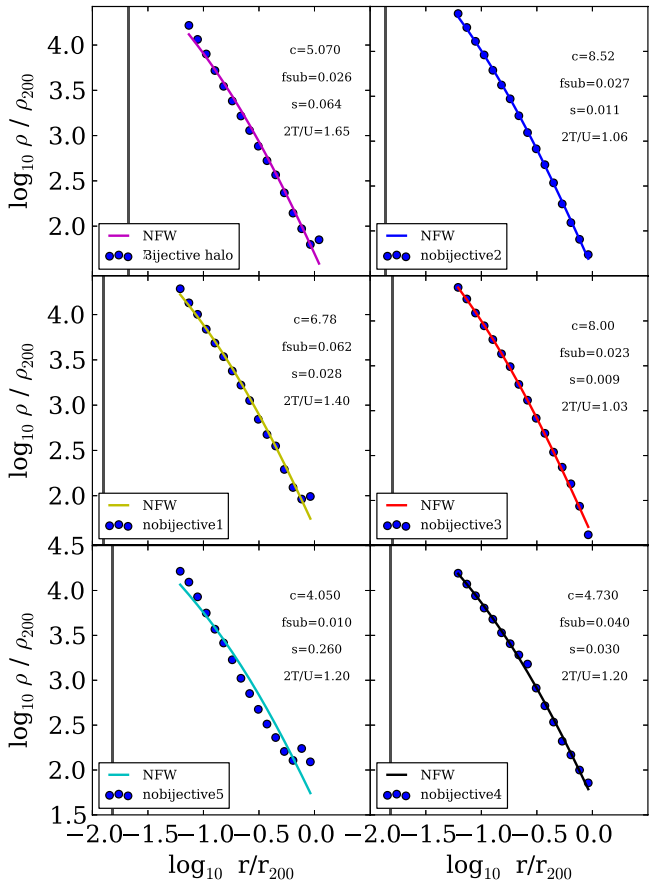

Figure 11. Density profiles, $\rho(r)$, for each of the Dhaloes shown in Fig. 9. The colour of the fitted NFW curve matches the colour coding of the individual Dhaloes in Fig. 9. The two-parameter, δ_c and r_s, NFW least-square fits were performed over the radial range $0.05 < r/r_{200} < 1$, shown by the black circles in Fig. 9. The minimum fit radius $r/r_{200} = 0.05$ is always larger than the convergence radius derived by Power et al. (2003), which we indicate by the solid vertical line in each panel.

mass is calculated using all the particles within r_{200}, not only those belonging to the FoF or Dhalo. We require $s < 0.07$ for relaxed haloes.

(iii) The virial ratio, $2T/|U|$, where T is the total kinetic energy of halo particles within r_{200} and U is their gravitational potential self-energy. We require $2T/|U| < 1.35$ for our relaxed haloes. (For haloes with more than 5000 particles, we use a random subset of 5000 particles to estimate U.)

Fig. 9 shows a single FoF halo and its component Dhaloes which we use to illustrate the application of these selection criteria and ability of NFW profiles to fit secondary/non-bijective Dhaloes. The spherically averaged density profiles and our NFW fits to each of these Dhaloes are shown in Fig. 11 along with the values of the three selection parameters $f_{\rm sub}$, s and $2T/|U|$. The top-left panel of Fig. 11 shows the density profile and NFW fit for the main component of the FoF halo, which can be identified by the cyan circle in Fig. 9 which marks twice the half-mass radius of the most massive substructure in the FoF halo. In previous analyses of FoF haloes, such as Neto et al. (2007), this would be the only density profile fitted to the mass distribution shown in Fig. 9. The bijectively matched Dhalo has the same centre as the FoF halo and the NFW fit is performed on all the mass within r_{200} (indicated by the concentric black circle); consequently, the density profile and NFW fit of the bijectively matched Dhalo are necessarily identical to that of the corresponding FoF halo. Examining this region in Fig. 9, we can clearly see that the mass distribution is asymmetric and has several distinct substructures indicative of a recent merger. This halo is not relaxed according to the above selection criteria as it fails to satisfy the cut on $2T/|U|$. Also its value of the centre offset, s, comes close to the threshold. The NFW fit to its density profile can be seen to have significant deviations at both large and small radii.

We are also interested in whether NFW profiles provide acceptable fits to the other Dhaloes found within this single FoF halo. These are shown in the remaining panels of Fig. 11. According to the selection criteria, three of these Dhaloes (those in the right-hand column) are relaxed. These are the blue, red and black Dhaloes in Fig. 9 and their density profiles are shown, respectively, in the top, middle and bottom-right panels of Fig. 11. In all cases, we see that the NFW fits provide a good description of the mass profile of these relaxed Dhaloes. The remaining two Dhaloes fail one or other of the selection criteria. The yellow Dhalo of Fig. 9, whose density profile is shown in the middle-left panel of Fig. 11, marginally fails the cut on $2T/|U|$. The cyan Dhalo of Fig. 9, whose density profile is shown in the bottom-left panel of Fig. 11, which strongly exceeds the threshold on s, can be seen to be very poorly fitted by the NFW profile and have a particularly low concentration. This Dhalo is very close to being within twice the half-mass radius of the most massive substructure of the FoF halo, marked by the cyan circle in Fig. 9. This being the radius used by the Dhalo algorithm as part of its criteria to determine whether two subhaloes should be considered as two distinct haloes or components of the same halo. It is this proximity to a merger that both creates the large offset, s, between the potential minimum and the centre of mass within r_{200} and distorts the object's density profile. We also note that this Dhalo has the most extreme ratio of r_{200} to twice its half-mass radius. In Fig. 4, we saw that for isolated haloes r_{200} and twice the half-mass radius were very comparable, but in contrast we see in Fig. 9 that the r_{200} of secondary Dhaloes can be significantly boosted by the density of the surrounding environment.

This systematic difference in the ratio of Dhalo mass to M_{200} for bijective and non-bijective Dhaloes is illustrated in Fig. 8 which should be contrasted with the right-hand panel of Fig. 2. We see that the scatter in the ratio of $M_{\rm Dhalo}/M_{200}$ is considerably larger for the non-bijective Dhaloes than it is for bijective Dhaloes. For bijective Dhaloes, the 5–95 per cent range of the distribution spans only a 30 per cent range in the ratio $M_{\rm Dhalo}/M_{200}$, while this is increased to approximately a factor of 2 for the non-bijective Dhaloes. In addition, the median $M_{\rm Dhalo}/M_{200}$ ratio is reduced from 1.2 for bijective Dhaloes to $\approx$0.95 for non-bijective Dhaloes. These differences are principally caused by the way the SUBFIND algorithm (Springel et al. 2001) is affected by the local environment. SUBFIND locates the edge of a substructure by searching for a saddle point in the density distribution. Hence, if the same substructure is placed in a denser environment, this will move the saddle point in and reduce the mass that SUBFIND associates with the substructure (see Muldrew, Pearce & Power 2011 for a detailed discussion). As a Dhalo mass is simply the sum of the masses of the subhaloes from which it is composed, this in turn reduces the mass assigned to the Dhalo. This systematic dependence of Dhalo mass on environment is one of the reasons why instead of directly using the Dhalo mass as input to GALFORM semi-analytic model we force the halo masses in the halo merger trees to increase monotonically so that they do not artificially decrease, just prior to mergers, due to such environmental effects.

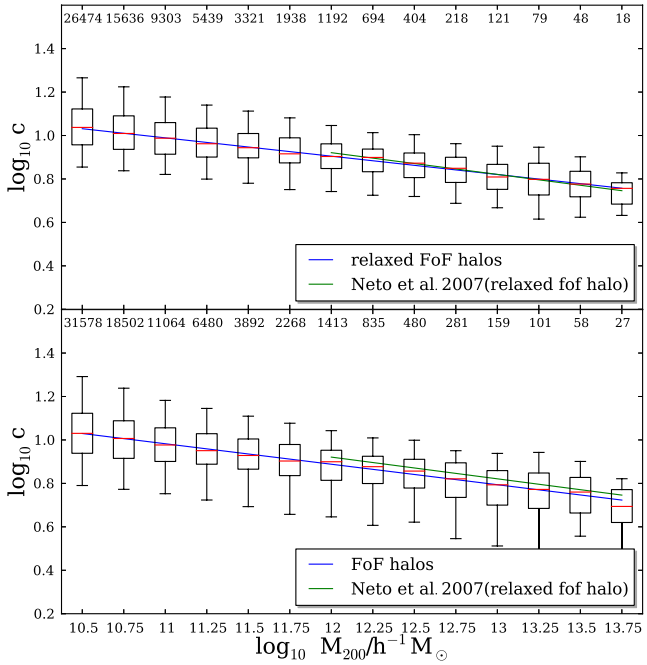

Figure 12. The mass–concentration relation for relaxed FoF haloes in MSII (top panel) and for all the FoF haloes (bottom panel). The boxes represent the 25 per cent and 75 per cent centiles of the distribution, while the whiskers show the 5 per cent and 95 per cent tails. The numbers on the top of each panel indicate the number of haloes in each mass bin. The median concentration as a function of mass is shown by the blue solid line and is well fitted by the linear relations given in equations (5) and (6). The green lines in each panel correspond to fits of Neto et al. (2007).

4.3 The mass–concentration relation

Here we compare the mass–concentration relation for FoF haloes that we find in the high-resolution MSII simulation with that found by Neto et al. (2007) in the lower resolution Millennium Simulation.[5] We then go on to compare this relation with the relation we find for the secondary/non-bijective Dhaloes. There is no need to separately look at the bijective Dhaloes as their M_{200} and c are necessarily the same as that of the corresponding FoF haloes as they have the same centre and all the surrounding mass is used in the fit. As in Neto et al. (2007), the mass we use in these relations is the M_{200} of the NFW fit rather than the directly measured value. Fig. 12 shows concentration as a function of mass for the range $10^{10.5} < M_{200}/h^{-1}\,\mathrm{M}_\odot < 10^{13.75}$ for our catalogue of FoF haloes. The top panel is for our *relaxed* FoF halo sample, while the bottom panel shows results for all the FoF haloes, including systems that do not meet our equilibrium criteria. In each case, we find a significant spread in concentration at fixed mass with a weak trend for decreasing concentration with increasing mass. This is generally interpreted (Navarro et al. 1995, 1996, 1997; Bullock et al. 2001; Eke et al. 2001; Neto et al. 2007; Gao et al. 2008) as reflecting the typical formation time of the halo with the lowest mass haloes forming earliest and having high-density cores which reflect the density of the universe at the time they formed. The dependence of the median concentration of FoF haloes on mass is well described by the power-law fit

$$c_{200} = 5.45\left(M_{200}/10^{14}\,h^{-1}\,\mathrm{M}_\odot\right)^{-0.084}, \quad (5)$$

for *relaxed* haloes and by

$$c_{200} = 5.01\left(M_{200}/10^{14}\,h^{-1}\,\mathrm{M}_\odot\right)^{-0.094} \quad (6)$$

for all haloes. These fits were performed only over the mass range $10^{10.5} < M_{200}/h^{-1}\,\mathrm{M}_\odot < 10^{13.75}$ due to poor statistics at higher masses and are shown by the blue solid lines in Fig. 12. Also shown in Fig. 12 is the fit for the median concentration for relaxed haloes found by Neto et al. (2007). We plot these green lines only for $M_{200} > 10^{12}/h^{-1}\,\mathrm{M}_\odot$ corresponding to the resolution limit of their study. We see that over the overlapping mass range, our median concentrations agree very well with those of Neto et al. (2007) indicating that the mass profiles over the fitted radial range, $-2.5 < \log(r/r_{200}) < 0$, are not affected by mass resolution. Our fit is also similar to the relation $c_{200} = 5.6(M_{200}/10^{14}\,h^{-1}\,\mathrm{M}_\odot)^{-0.098}$ found by Macciò et al. (2007) for relaxed haloes. The small difference could be because they fit the mean rather than median of the relation or due to differences in the criteria used to select relaxed haloes. Like us and Neto et al. (2007), Macciò et al. (2007) find that unrelaxed haloes have systematically lower concentrations.

Having demonstrated that for FoF haloes we recover a mass–concentration relation which is in very accurate agreement with previous work (Macciò et al. 2007; Neto et al. 2007), we now want to compare mass–concentration relations for our bijective and non-bijective Dhaloes. The mass–concentration relation we find for the bijective Dhaloes is practically identical to that of the FoF haloes plotted in Fig. 12 and so we have chosen not to effectively repeat the same plot. The similarity is inevitable as Fig. 1 shows that for masses greater than $10^{10.5}\,h^{-1}\,\mathrm{M}_\odot$, for which we can measure concentrations, the fraction of FoF haloes that have bijective matches with Dhaloes is greater than 95 per cent and these bijectively matched haloes have identical centres and so identical fitted NFW mass profiles.

In Fig. 13, we show the mass–concentration for relaxed and all non-bijective Dhaloes. These haloes are all secondary fragments of FoF haloes and so are a completely disjoint catalogue of haloes to those represented in the FoF mass–concentration relations of Fig. 12. To aid in comparing the two sets of relations, we plot the power-law fits to the median mass–concentration relations of Fig. 12 as dashed lines in Fig. 13. It can be seen that these are very similar to the power-law fits to the median relations

$$c_{200} = 4.90\left(M_{200}/10^{14}\,h^{-1}\,\mathrm{M}_\odot\right)^{-0.093}, \quad (7)$$

for *relaxed* and

$$c_{200} = 5.01\left(M_{200}/10^{14}\,h^{-1}\,\mathrm{M}_\odot\right)^{-0.095} \quad (8)$$

for all the non-bijective Dhaloes which are shown by the solid lines in Fig. 13.

Comparison of the bars and whiskers in Figs 12 and 13 shows that the not only do the median mass–concentration relations for FoF and non-bijective Dhaloes agree very well, but the distributions of concentrations about the medians are also quite similar. The large number of haloes we have in the MII simulation enables us to look at these distributions in more detail and in Fig. 14 we show

[5] As a precise test of our methods, we first applied our analysis to FoF haloes in the milli-MSII simulation, which has the same volume, initial conditions and data format as MSII (Boylan-Kolchin et al. 2009), but lower mass resolution, equal to that of the Millennium Simulation (Springel et al. 2005) analysed by Neto et al. (2007). We found precise agreement with the mass–concentration relationship published in Neto et al. (2007).

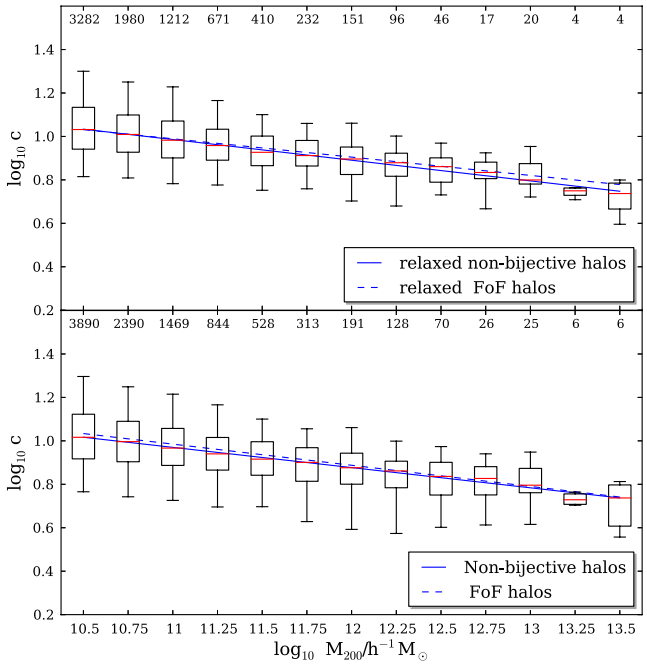

Figure 13. The mass–concentration relation for relaxed non-bijective Dhaloes in MSII (top panel) and for all the non-bijective Dhaloes (bottom panel). The boxes represent the 25 per cent and 75 per cent centiles of the distribution, while the whiskers show the 5 per cent and 95 per cent tails. The numbers on the top of each panel indicate the number of haloes in each mass bin. The median concentration as a function of mass is shown by the blue solid line and is well fitted by the linear relations given in equations (7) and (8). The blue dashed line in each panel repeats the fits to the median mass–concentration relation for FoF haloes shown in Fig. 12.

histograms of the concentration, distributions along with log-normal approximations

$$P(\log_{10} c) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{\log_{10} c - \langle \log_{10} c \rangle}{\sigma}\right)^2\right], \qquad (9)$$

for two mass bins centred on 10^{11} and 10^{12} h^{-1} $M_\odot$. We see in all cases that the non-bijective Dhaloes have a very similar distribution of concentrations as the distribution of the corresponding FoF sample and that both are approximated accurately by log-normal distributions. Note that in both cases we are binning haloes by the M_{200} of their fitted NFW profile and so we are affected by the Dhalo mass being perturbed and suppressed in non-bijective Dhaloes. We recall that the FoF sample is essentially the same as the sample of bijectively matched Dhaloes and so we conclude that concentration distribution is essentially the same for both the primary Dhaloes and those that are secondary fragments of FoF haloes. In all cases, the concentration distributions for the relaxed samples have slightly higher median concentrations and smaller dispersions than the corresponding complete mass selected samples.

Also of interest is the fraction of both FoF haloes and non-bijective Dhaloes that satisfy the equilibrium criteria. From the number of objects per mass bin given in the labels in Figs 12 and 13, this can be seen to be in the range of 80–85 per cent for both FoF and Dhaloes. One might at first expect that many multinucleated FoF haloes would fail both the threshold on the asymmetry, s, and the fraction of mass in substructures, $f_{\rm sub}$. However, as these statistics are evaluated only using the mass within r_{200} and not across the whole FoF halo, $\gtrsim$98 per cent of FoF haloes pass the substructure threshold and $\gtrsim$88 per cent the asymmetry threshold. The first of these numbers is slightly lower for the non-bijective Dhaloes, i.e. only $\gtrsim$93 per cent pass the substructure threshold. However, those passing the more stringent asymmetry threshold is more comparable at $\gtrsim$86 per cent, while for both FoF and non-bijective Dhaloes $\gtrsim$93 per cent pass the criterion that the virial ratio $2T/|U| < 1.35$. Consequently, the fraction of the non-bijective Dhaloes that pass the relaxation criteria is very similar to that for the FoF or bijective Dhaloes. Hence, in both cases the mass–concentration distributions that we have quantified are representative of the vast majority of the haloes.

5 CONCLUSIONS

We have used the high-resolution Millennium Simulation II cosmological N-body simulation to quantify the properties of haloes defined by the Dhalo algorithm. This algorithm is designed to produce merger trees suitable for use with the semi-analytic galaxy formation model, GALFORM. We have included a full description of the Dhalo algorithm which produces a set of haloes, and the merger trees that describe their hierarchical evolution, that are consistent between subsequent snapshots of the N-body simulations. We have presented the properties of the Dhaloes by comparing them with the corresponding properties of the much more commonly used FoF haloes (Davis et al. 1985).

We have shown that unlike the FoF algorithm the Dhalo algorithm is successful in avoiding distinct masses and concentrations being prematurely linked together into a single halo when their diffuse outer haloes touch. We have also illustrated how some Dhaloes can be composed of more than one FoF halo. This occurs as structure formation in CDM models is not strictly hierarchical and occasionally a halo, after falling into a more massive halo, may escape to beyond the virial radius of the more massive halo. For the purposes of the GALFORM semi-analytic model, it is convenient to consider such haloes as remaining as satellites of the main halo. We find that such remerged FoF haloes are not uncommon, but contribute very little mass to the larger haloes to which they are (re)attached.

Despite the complex mapping between FoF and Dhaloes, which results in a significant fraction of FoF haloes being broken up into multiple Dhaloes while other FoF haloes get (re)merged into a single Dhalo, we find that the overall mass functions of the two sets of haloes are very similar. The mass functions of our Dhalo and FoF halo catalogues are both reasonably well fitted over the mass range of 10^8–$10^{13.5}$ h^{-1} $M_\odot$ by currently popular analytic mass functions such as those of Warren et al. (2006) and Reed et al. (2007).

Approximately 90 per cent of the Dhaloes have a unique one-to-one, bijective, match with a corresponding FoF halo. For this subset of haloes, the mass of the Dhalo, $M_{\rm Dhalo}$, correlates much more closely with the standard virial mass, M_{200}, than does the FoF mass. The median $M_{\rm FoF}/M_{200} = 1.2$ and 90 per cent of the distribution of this mass ratio span a factor 1.9, while for the same Dhaloes the median $M_{\rm Dhalo}/M_{200} = 1.15$ and corresponding width of the distribution span only a factor 1.3. The larger scatter in the FoF case is often caused by secondary mass concentrations that lie outside the r_{200} radius of the main substructure and are linked into the FoF halo by particle bridges in overlapping diffuse haloes. The non-bijective Dhaloes have a wider distribution, with 90 per cent of the distribution spanning a factor 2.2 and with the median ratio reduced to $M_{\rm Dhalo}/M_{200} = 0.95$. This is due to the SUBFIND substructure finder, which is part of the Dhalo algorithm, assigning less mass to subhaloes when they move into overdense environments. When utilized in GALFORM, this systematic loss of mass is not an issue

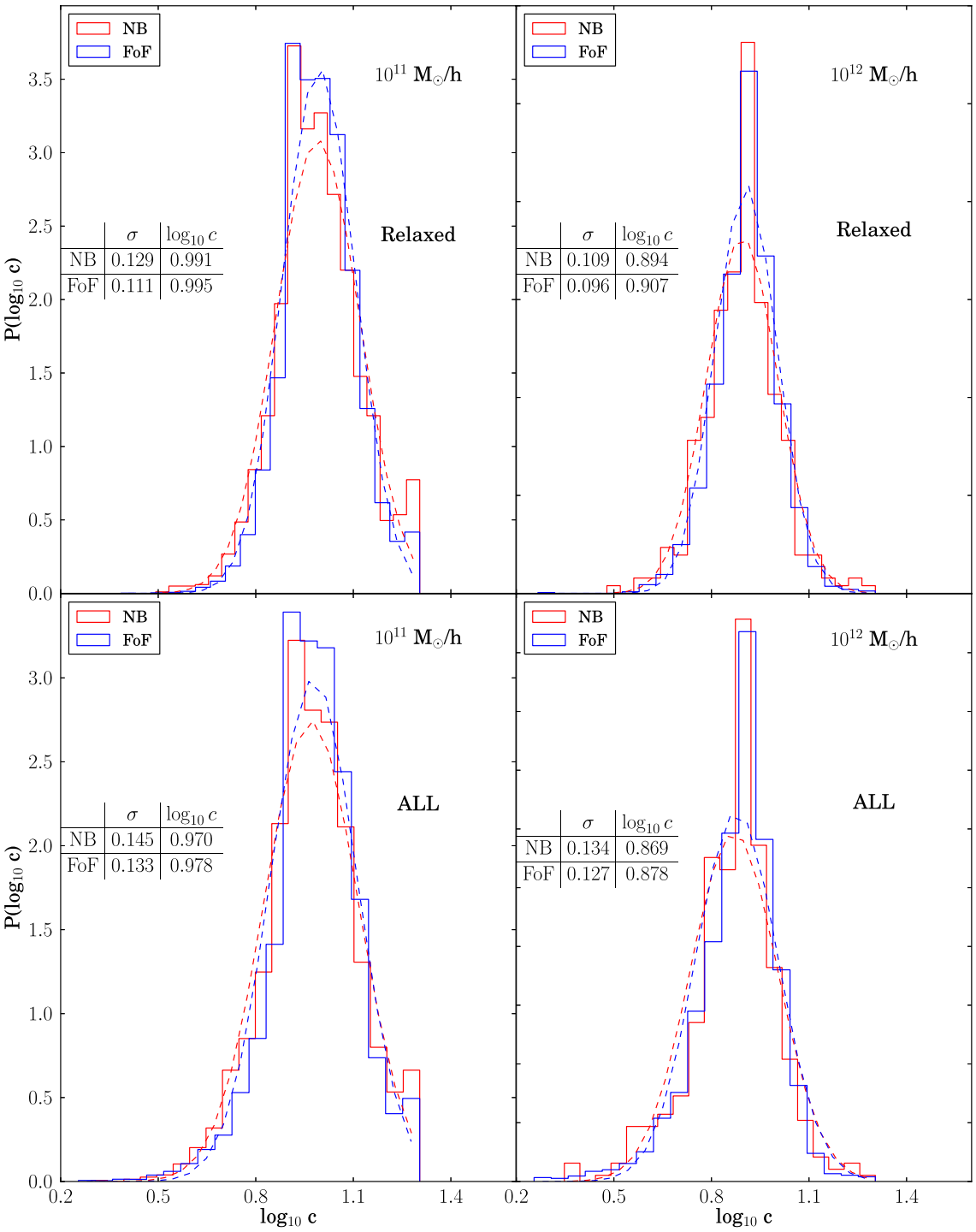

Figure 14. The distribution of concentrations for haloes in the two mass bins $10.75 < \log_{10} M_{200}/h^{-1}\mathrm{M}_\odot < 11.25$ and $11.75 < \log_{10} M_{200}/h^{-1}\mathrm{M}_\odot < 12.25$. The upper panels are for samples of relaxed haloes while the bottom panels are for all haloes whether or not they satisfy the relaxation criteria. In each panel, the blue histogram is for FoF haloes and the red histogram is for Dhaloes that do not have bijective matches to FoF haloes. The smooth curves are log-normal approximations with the same $\log_{10} c$ and second moment, σ, as the measured distributions. The corresponding values of $\log_{10} c$ and σ are given in the legend.

as the merger trees are preprocessed and mass is added back in to ensure that the branches of the GALFORM merger trees always have monotonically increasing masses.

The high resolution of the MSII Simulation has allowed us to study the density profiles and concentrations of both FoF and Dhaloes over a wide range of mass. To avoid contaminating our samples with unrelaxed haloes for which fitting smooth spherically symmetric profiles is inappropriate, we exclude unrelaxed haloes using the relaxation criteria from Neto et al. (2007). We find that 80 per cent of both FoF and Dhaloes are relaxed according to these criteria. For FoF haloes, we accurately reproduce the mass–concentration distribution found by Neto et al. (2007) at high masses and extend the distribution to much lower masses. Combining our results with those of Macciò et al. (2007) and Neto et al. (2007),

we find that a single power law reproduces the mass–concentration relation for over five decades in mass. We also find that the mass–concentration distributions for Dhaloes agree very accurately with those for FoF haloes. This is true even for non-bijective Dhaloes which are secondary components of FoF haloes. The properties of such haloes have generally been overlooked in previous studies. We show that the distributions of concentrations around the mean mass–concentration relation are well described by log-normal distributions for both the FoF and Dhaloes.

ACKNOWLEDGEMENTS

This work was supported by the Science and Technology Facilities (grant number ST/F001166/1). LJ acknowledges the support of a Durham Doctoral Studentship. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

Audit E., Teyssier R., Alimi J., 1998, A&A, 333, 779
Baugh C. M., Lacey C. G., Frenk C. S., Granato G. L., Silva L., Bressan A., Benson A. J., Cole S., 2005, MNRAS, 356, 1191
Behroozi P., Wechsler R., Wu H.-Y., 2012, Astrophysics Source Code Library, record ascl:1210.008
Behroozi P. S., Wechsler R. H., Wu H.-Y., Busha M. T., Klypin A. A., Primack J. R., 2013, ApJ, 763, 18
Benson A. J., Bower R., 2010, MNRAS, 405, 1573
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38
Berlind A. A., Weinberg D. H., 2002, ApJ, 575, 587
Berlind A. A. et al., 2003, ApJ, 593, 1
Bhattacharya S., Heitmann K., White M., Lukić Z., Wagner C., Habib S., 2011, ApJ, 732, 122
Binney J., 1977, ApJ, 215, 483
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, MNRAS, 398, 1150
Bullock J. S., Kolatt T. S., Sigad Y., Somerville R. S., Kravtsov A. V., Klypin A. A., Primack J. R., Dekel A., 2001, MNRAS, 321, 559
Bundy K., Ellis R. S., Conselice C. J., 2005, ApJ, 625, 621
Busha M. T., Alvarez M. A., Wechsler R. H., Abel T., Strigari L. E., 2010, ApJ, 710, 408
Cole S., 1991, ApJ, 367, 45
Cole S., Lacey C., 1996, MNRAS, 281, 716
Cole S., Aragon-Salamanca A., Frenk C. S., Navarro J. F., Zepf S. E., 1994, MNRAS, 271, 781
Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, MNRAS, 319, 168
Colless M. et al., 2001, MNRAS, 328, 1039
Courtin J., Rasera Y., Alimi J., Corasaniti P., Boucher V., Füzfa A., 2011, MNRAS, 410, 1911
Creasey P., Theuns T., Bower R. G., Lacey C. G., 2011, MNRAS, 415, 3706
Crocce M., Fosalba P., Castander F. J., Gaztañaga E., 2010, MNRAS, 403, 1353
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Diemand J., Kuhlen M., Madau P., 2006, ApJ, 649, 1
Einasto J., Klypin A. A., Saar E., Shandarin S. F., 1984, MNRAS, 206, 529
Eke V. R., Navarro J. F., Steinmetz M., 2001, ApJ, 554, 114
Eke V. R. et al., 2004, MNRAS, 355, 769
Evrard A. E. et al., 2002, ApJ, 573, 7
Evrard A. E. et al., 2008, ApJ, 672, 122
Font A. S. et al., 2008, MNRAS, 389, 1619
Font A. S. et al., 2011, MNRAS, 417, 1260
Frenk C. S., White S. D. M., Davis M., Efstathiou G., 1988, ApJ, 327, 507
Gao L., Navarro J. F., Cole S., Frenk C. S., White S. D. M., Springel V., Jenkins A., Neto A. F., 2008, MNRAS, 387, 536
Gill S. P. D., Knebe A., Gibson B. K., 2005, MNRAS, 356, 1327
Giocoli C., Pieri L., Tormen G., 2008, MNRAS, 387, 689
Gottlöber S., Yepes G., 2007, ApJ, 664, 117
Green A. M., Hofmann S., Schwarz D. J., 2004, MNRAS, 353, L23
Guo Q. et al., 2011, MNRAS, 413, 101
Han J., Jing Y. P., Wang H., Wang W., 2012, MNRAS, 427, 2437
Harker G., Cole S., Helly J., Frenk C., Jenkins A., 2006, MNRAS, 367, 1039
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Heitmann K., Lukić Z., Habib S., Ricker P. M., 2006, ApJ, 642, L85
Helly J. C., Cole S., Frenk C. S., Baugh C. M., Benson A., Lacey C., 2003, MNRAS, 338, 903
Huchra J. P., Geller M. J., 1982, ApJ, 257, 423
Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, MNRAS, 321, 372
Jiang F., van den Bosch F. C., 2014, MNRAS, preprint (arXiv:1311.5225)
Kauffmann G., White S., 1993, MNRAS, 261, 921
Kim H.-S., Baugh C. M., Benson A. J., Cole S., Frenk C. S., Lacey C. G., Power C., Schneider M., 2011, MNRAS, 414, 2367
Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 516, 530
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Knebe A. et al., 2011, MNRAS, 415, 2293
Knebe A. et al., 2013, MNRAS, 435, 1618
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Koposov S. E., Yoo J., Rix H.-W., Weinberg D. H., Macciò A. V., Escudé J. M., 2009, ApJ, 696, 2179
Lacey C., Cole S., 1994, MNRAS, 271, 676
Lacey C., Silk J., 1991, ApJ, 381, 14
Lagos C. D. P., Baugh C. M., Lacey C. G., Benson A. J., Kim H.-S., Power C., 2011, MNRAS, 418, 1649
Lemson G., the Virgo Consortium, 2006, preprint (astro-ph/0608019)
Linder E. V., Jenkins A., 2003, MNRAS, 346, 573
Lokas E. L., Bode P., Hoffmann Y., 2004, MNRAS, 349, 595
Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, ApJ, 692, 931
Lukic Z., Reed D., Habib S., Heitmann K., 2009, ApJ, 692, 217
Macciò A. V., Dutton A. A., van den Bosch F. C., Moore B., Potter D., Stadel J., 2007, MNRAS, 378, 55
Macciò A. V., Dutton A. A., van den Bosch F. C., 2008, MNRAS, 391, 1940
Macciò A. V., Kang X., Fontanot F., Somerville R. S., Koposov S., Monaco P., 2010, MNRAS, 402, 1995
Merson A. I. et al., 2013, MNRAS, 429, 556
Moreno J., Giocoli C., Sheth R. K., 2008, MNRAS, 391, 1729
Muldrew S. I., Pearce F. R., Power C., 2011, MNRAS, 410, 2617
Muñoz J. A., Madau P., Loeb A., Diemand J., 2009, MNRAS, 400, 1593
Murray S., Power C., Robotham A., 2013a, Astron. Comput., 3, 23
Murray S. G., Power C., Robotham A. S. G., 2013b, MNRAS, 434, L61
Navarro J. F., Frenk C. S., White S. D. M., 1995, MNRAS, 275, 720
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493
Neistein E., Dekel A., 2008, MNRAS, 388, 1792
Neto A. et al., 2007, MNRAS, 381, 1450
Onions J. et al., 2012, MNRAS, 423, 1200
Parkinson H., Cole S., Helly J., 2008, MNRAS, 383, 557
Peacock J. A., Smith R. E., 2000, MNRAS, 318, 1144
Percival W. J. et al., 2001, MNRAS, 327, 1297
Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, MNRAS, 338, 14
Power C., Knebe A., Knollmann S. R., 2012, MNRAS, 419, 1576

Press W. H., Davis M., 1982, ApJ, 259, 449
Reed D., Gardner J., Quinn T., Stadel J., Fardal M., Lake G., Fabio G., 2003, MNRAS, 346, 565
Reed D., Bower R., Frenk C. S., Jenkins A., Theuns T., 2007, MNRAS, 374, 2
Rees M. J., Ostriker J. P., 1977, MNRAS, 179, 541
Sanchez A. G., Baugh C. M., Percival W. J., Peacock J. K., Padilla N. D., Cole S., Frenk C. S., Norberg P., 2006, MNRAS, 366, 1
Schaye J. et al., 2010, MNRAS, 402, 1536
Seljak U., 2000, MNRAS, 318, 203
Sheth R. K., Tormen G., 1999, MNRAS, 308, 119
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481
Spergel D. N. et al., 2003, ApJS, 148, 175
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001, MNRAS, 328, 726
Springel V. et al., 2005, Nature, 435, 629
Springel V. et al., 2008, MNRAS, 391, 1685
Srisawat C. et al., 2013, MNRAS, 436, 150
Summers F. J., Davis M., Evrard A. E., 1995, ApJ, 454, 1
Thomas P. A., Muanwong O., Pearce F. R., Couchman H. M. P., Edge A. C., Jenkins A., Onuora L., 2001, MNRAS, 324, 450
Tinker J., Kravtsov A. V., Klypin A., Abazajian K., Warren M., Yepes G., Gottlöber S., Holz D. E., 2008, ApJ, 688, 709
van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029
Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, ApJ, 646, 881
Watson W. A., Iliev I. T., D'Aloisio A., Knebe A., Shapiro P. R., Yepes G., 2013, MNRAS, 433, 1230
White M., 2001, ApJS, 143, 241
White S. D. M., Frenk C. S., 1991, ApJ, 379, 52
White S. D. M., Rees M. J., 1978, MNRAS, 183, 341

APPENDIX A: CONSTRUCTING DHALO MERGER TREES

Here we describe in detail the algorithm used to produce the Dhalo merger trees. These merger trees are intended to be used as input to the GALFORM semi-analytic model of galaxy formation. The need for consistency between the halo model used in the semi-analytic calculation and the *N*-body simulation imposes some requirements on the construction of the merger trees.

The GALFORM galaxy formation model makes the approximation that mergers between haloes are instantaneous events and assumes that haloes, once merged, do not fragment. However, in *N*-body simulations halo mergers take a finite amount of time and it is not uncommon for a halo falling into another, more massive halo to escape to well beyond the virial radius after its initial infall (Gill, Knebe & Gibson 2005; Ludlow et al. 2009). We therefore need to choose when to consider *N*-body haloes to have merged in the semi-analytic model and define our haloes such that they remain merged at all later times. We also wish to define the haloes used to construct the trees such that, as far as possible, they resemble the spherically symmetric, virialized objects assumed in the galaxy formation model. Quantifying the extent to which we have achieved this is one of the main aims of this paper.

A1 Halo catalogues

The first step in building the merger trees is to use the FoF (Davis et al. 1985) and SUBFIND algorithms (Springel et al. 2001) to identify haloes and subhaloes in all of the simulation snapshots. The SUBFIND algorithm decomposes each FoF halo into subhaloes by identifying self-bound density maxima. Usually the most massive subhalo contains most of the mass of the original FoF halo. Secondary density maxima give rise to additional subhaloes. Compared to the FoF halo the most massive subhalo does not include any of the mass assigned to other subhaloes (a simulation particle can only belong to one subhalo) nor does it include particles that are not gravitationally bound to it. Some of the lowest mass FoF haloes have no self-bound subhaloes and most FoF haloes have at least some 'fuzz' of unbound particles which belong to no subhalo. FoF haloes with no self-bound subhaloes are not used in the construction of the merger trees.

A2 Building the subhalo merger trees

Before we can construct the Dhalo merger trees, it is necessary to define subhalo merger trees by identifying the descendant of each subhalo. The code we use to do this was included in the merger trees comparison project carried out by Srisawat et al. (2013) under the name D-TREES. The project concluded that it was a desirable feature for a merger tree code to use particle IDs to match haloes between snapshots and have the ability to search multiple snapshots for descendants. The latter requirement was due to the tendency of the Amiga's Halo Finder (Knollmann & Knebe 2009) used in the project to temporarily fail to detect substructures during mergers.

Since SUBFIND suffers from a similar problem, we allow for the possibility that the descendant of a subhalo may be found more than one snapshot later. Our approach is to devise an algorithm which can identify the descendant of a halo at any single, later snapshot, apply it to the next $N_{\rm step}$ snapshots (where $N_{\rm step} = 5$) and pick one of these $N_{\rm step}$ possible descendants to use as the descendant of the subhalo in the merger trees.

Alternative solutions to this problem include allowing the merger tree code to modify the subhalo catalogue to ensure consistency of subhalo properties between snapshots (CONSISTENTTREES; Behroozi et al. 2013) and using information from previous snapshots to define the subhalo catalogue (HBT; Han et al. 2012).

In common with all but one of the merger tree codes in the comparison (JMERGE, which relies entirely on aggregate properties of the haloes), we identify descendants by finding subhaloes at different snapshots which have particles in common.

A2.1 Identifying a descendant at a single, later snapshot

To find the descendant at snapshot j, of a halo which exists at an earlier snapshot, i, the following method is used. For each halo containing N_p particles, the $N_{\rm link}$ most bound are identified, where $N_{\rm link}$ is given by

$$N_{\rm link} = \min(N_{\rm linkmax}, \max(f_{\rm trace} N_p, N_{\rm linkmin})) \quad \text{(A1)}$$

with $N_{\rm linkmin} = 10$, $N_{\rm linkmax} = 100$ and $f_{\rm trace} = 0.1$.

For each of the haloes at snapshot i, descendant candidates are found by locating all haloes at snapshot j which received at least one particle from the earlier halo. Then, a single descendant is chosen from these candidates as follows. If any of the descendant candidates received a larger fraction of their $N_{\rm link}$ most bound particles from the progenitor halo than from any other halo at the earlier snapshot, then the descendant is chosen from these candidates only and the halo at snapshot i will be designated the main progenitor of the chosen descendant; otherwise, all candidates are considered and the halo will not be the main progenitor of its descendant. The descendant of the halo at snapshot i is taken to be the remaining candidate which received the largest fraction of the $N_{\rm link}$ most bound of the progenitor halo. For each halo at snapshot j, this method identifies

zero or more progenitors of which at most one may be a main progenitor. Note that it is not guaranteed that a main progenitor will be found for every halo.

By following the most bound part of the subhalo, we ensure that if the core of a subhalo survives at the later snapshot, it is identified as the descendant irrespective of how much mass has been lost. It also means that in the cases where an object at the later snapshot has multiple progenitors, we can determine which one of the progenitors contributed the largest fraction of the most bound core of the descendant object. We consider this main progenitor to have survived the merger while the other progenitors have merged on to it and ceased to exist as independent objects.

Fig. A1 shows three examples of this linking procedure. In the simplest case (left), a single, isolated subhalo B at snapshot j is identified as the descendant of subhalo A which exists at the earlier snapshot i. Since more of the most bound particles of subhalo B come from subhalo A than from any other subhalo, we conclude that A is the main progenitor of B. In the second case (centre), two subhaloes A and B merge to form subhalo C at the later snapshot. Subhalo A is determined to be the main progenitor because it contributed the largest fraction of the most bound particles of the descendant, C. In the third example (right), a satellite subhalo A exists within a more massive host halo. In this case, particles from the subhalo A are split between subhalo B and the host halo C at the later snapshot. While a large fraction (or even the vast majority) of the particles from subhalo A may belong to the host halo at the later snapshot, we choose subhalo B as the descendant because its most bound part came from subhalo A.

A2.2 Searching multiple snapshots for descendants

If a subhalo is not found to be the main progenitor of its descendant, this may indicate that the subhalo has merged with another subhalo and no longer exists as an independent object. However, it is also possible that the substructure finder has simply failed to identify the object at the later snapshot because it is superimposed on the dense central parts of a larger subhalo. Typically, this phase lasts for a small fraction of the host halo dynamical time (Behroozi et al. 2013) which in turn is much shorter than the usual interval between the snapshots of cosmological N-body simulations. Hence, by looking one snapshot ahead we will normally find the missed subhalo, but one can be unlucky and catch it half an orbit later when again it is hidden by the dense core of the more massive subhalo in which it is orbiting. Hence, looking several snapshots ahead exponentially suppresses this possibility. Thus, in order to distinguish between subhalo mergers and subhaloes which are just temporarily lost, it is necessary to search multiple snapshots for descendants.

In our algorithm for each snapshot i in the simulation, descendants are identified at later snapshots in the range $i + 1$ to $i + N_{\rm step}$ using the method described in Section A2.1. For each subhalo at snapshot i, this gives up to $N_{\rm step}$ possible descendants. One of these descendants is picked for use in the merger trees as follows: if the subhalo at snapshot i is the main progenitor of one or more of the descendants, the earliest of these descendants which does not have a main progenitor at a snapshot later than i is chosen. If no such descendant exists, the earliest descendant found is chosen irrespective of main progenitor status.

Descendants more than one snapshot later are only chosen in the cases where the earlier subhalo is the main progenitor – i.e. where the group still survives as an independent object. If the subhalo does not survive, we have no way to determine whether it merged immediately or if SUBFIND failed to detect it for one more snapshot prior to the merger, so we simply assume that the merger happened between snapshots i and $i + 1$.

Fig. A2 shows a case where a descendant more than one snapshot later is chosen. Subhalo A exists at snapshot i. Its descendant at snapshot $i + 1$ is found to be subhalo D. However, the most bound particles of D were not contributed by subhalo A, but by another

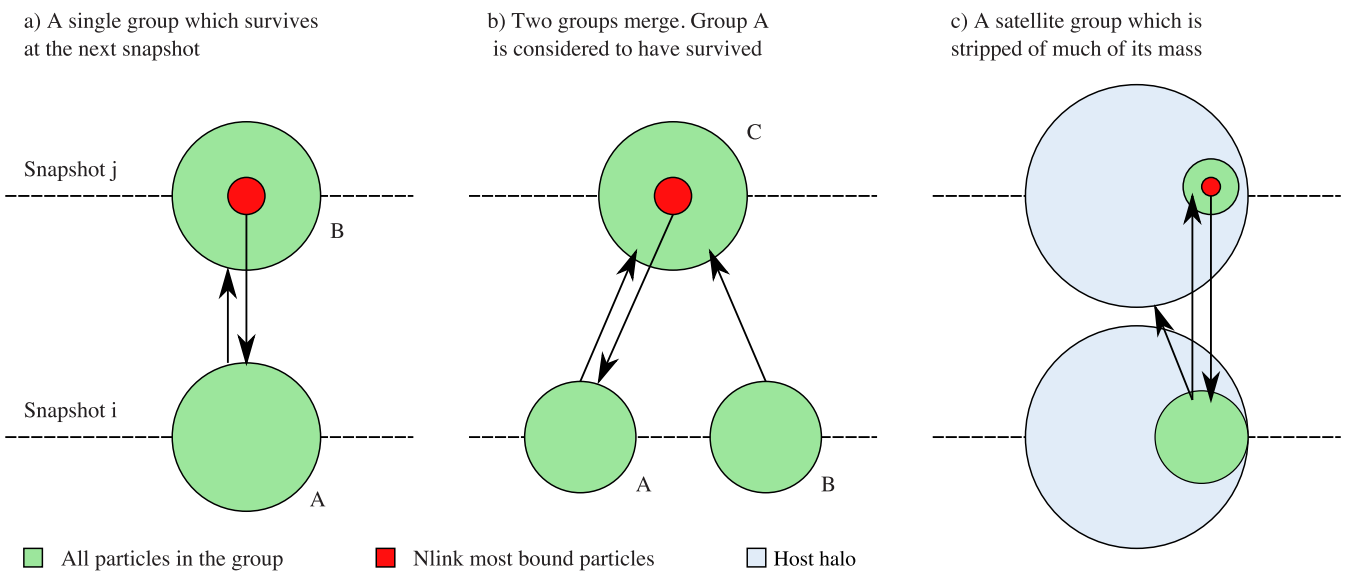

Figure A1. Schematic examples illustrating the method used to link SUBFIND subhaloes between pairs of snapshots i and j, where $i < j$. The green circles represent SUBFIND subhaloes. The most bound $N_{\rm link}$ particles in each subhalo at the later time are shown in red. From left to right are (a) a single, isolated subhalo which still exists at the next snapshot, (b) a merger between subhaloes A and B where more of the most bound particles of the merged halo C come from halo A than from any other halo and therefore halo A is considered to be the main progenitor of halo C, and (c) a satellite subhalo orbiting within a background halo which loses a large fraction of its particles to its host halo at the next snapshot but is still identified by SUBFIND. Arrows between green circles show the location of the majority of the particles in the subhalo at the later snapshot. Arrows starting from red circles show the location of the majority of the most bound particles at the earlier snapshot.

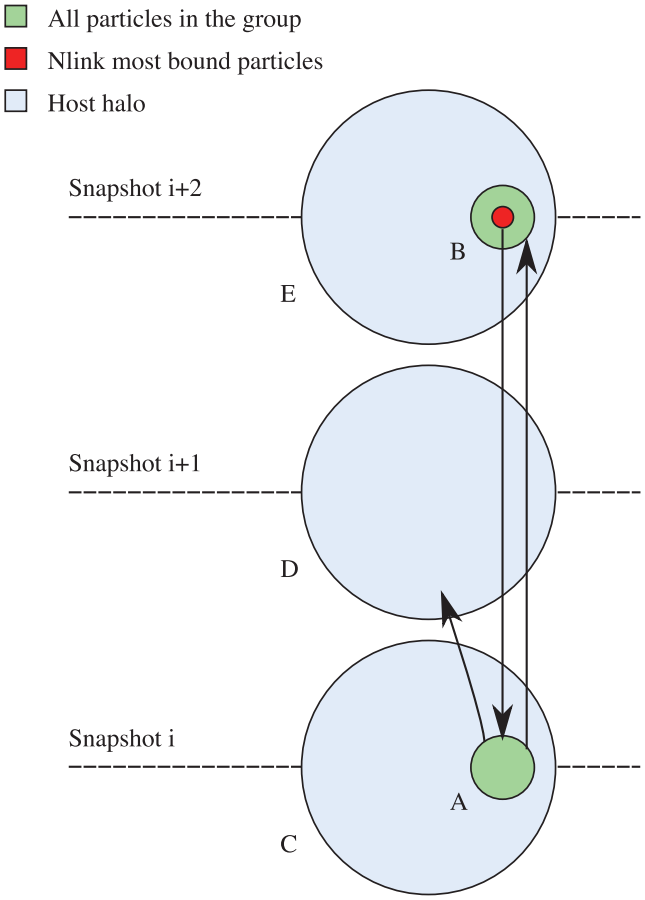

Figure A2. A schematic example of a case where the descendant of a subhalo is found to be more than one snapshot later. The green circles represent a satellite SUBFIND subhalo within a larger host halo which is represented by the blue circles. Three consecutive snapshots are shown.

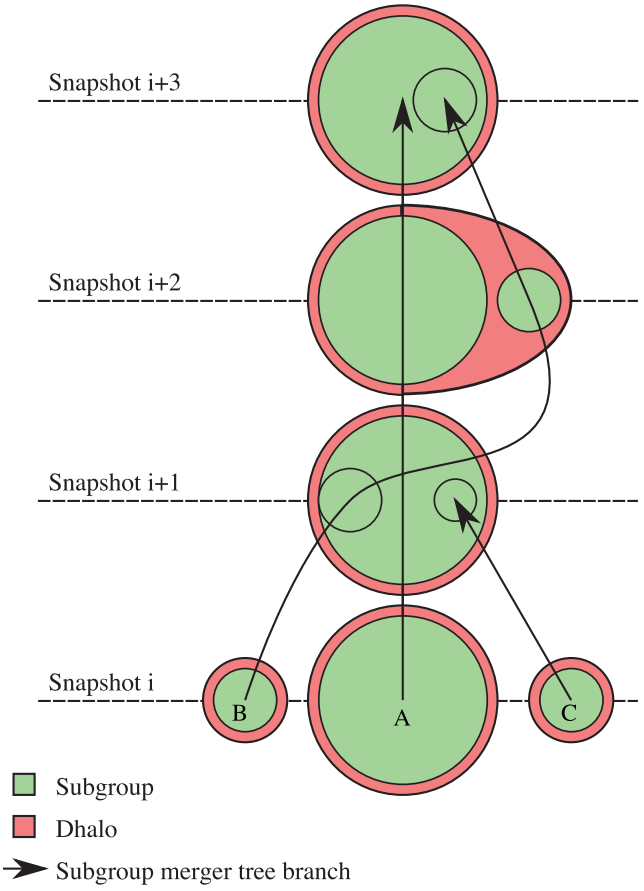

Figure A3. An example of a Dhalo merger tree showing two less massive haloes falling into another, more massive halo. Subhaloes are shown in green. Red areas indicate subhaloes which belong to the same Dhalo. The black arrows show branches of the subhalo merger tree.

progenitor, subhalo C. This means that A is not the main progenitor of its descendant at snapshot $i + 1$ and so it is necessary to consider possible descendants at later snapshots. Two subhaloes at snapshot $i + 2$ (B and E) receive particles from subhalo A. Since the most bound particles of subhalo B came from subhalo A, A is the main progenitor of B and subhalo B is taken to be the descendant of A.

A3 Constructing a halo catalogue

At this point we have a descendant for each subhalo. This is sufficient to define merger trees for the subhaloes. These SUBFIND trees can be split into 'branches' as follows. A new branch begins whenever a new subhalo forms (i.e. the subhalo has no progenitors). The remaining subhaloes that make up the branch are found by following the descendant pointers until either a subhalo is reached that is not the main progenitor of its descendant, a subhalo is reached that has no descendant or the final snapshot of the simulation is reached. Each of these branches represents the lifetime of an independent halo or subhalo in the simulation. We construct haloes and halo merger trees by grouping together these branches of the subhalo merger trees using methods which will be described below. We refer to the resulting collections of subhaloes as 'Dhaloes'. Fig. A3 shows an example of a Dhalo merger tree with the subhalo merger tree branches marked. In this case, there are three branches. Branch A is a single, massive halo which exists as an independent halo at all four snapshots. Branch B is a smaller halo which becomes a satellite subhalo within halo A, but continues to exist. Branch C is another small halo which briefly becomes a satellite before merging with A.

For each subhalo in an FoF halo, we identify the least massive, more massive 'enclosing' subhalo in the same FoF halo. Subhalo A is said to enclose subhalo B if B's centre lies within twice the half-mass radius of A. A pointer to the enclosing subhalo is stored for each subhalo that is enclosed. This produces a tree structure which is intended to represent the hierarchy of haloes, subhaloes, sub-subhaloes, etc. in the FoF halo. Any subhalo which is not enclosed by any other becomes a new Dhalo. Any subhaloes enclosed by this subhalo are assigned to the new Dhalo.

We then iterate through the snapshots from high redshift to low redshift. For each subhalo, we find the maximum number of particles it ever contained while it was the most massive subhalo in its parent FoF halo. If a satellite subhalo in a Dhalo retains a fraction $f_{\rm split}$ of its maximum isolated mass, then it is split from its parent Dhalo and becomes a new Dhalo. Any subhaloes enclosed by this subhalo are assigned to the new Dhalo too. We usually set $f_{\rm split} = 0.75$, so that when a halo falls into another, more massive halo, the two haloes will only be considered to have merged into one once the smaller halo has been stripped of some of its mass. This is to ensure that haloes artificially linked by the FoF algorithm are still treated as separate objects.

In some cases, a subhalo may escape from its parent halo. This happens to halo B in Fig. A3. For the purposes of semi-analytic galaxy formation modelling, we would like to continue to treat such

subhaloes as satellites in the parent halo so that each infalling halo contributes a single branch to the halo merger tree. This is done by merging such objects back on to the Dhalo they escaped from; the subhalo is recorded as a satellite within the original Dhalo at all later times regardless of its spatial position. Any subhaloes it encloses will also be considered to be part of this Dhalo.

In practice, the remerging is carried out in the following way. For each Dhalo *A*, we identify a descendant Dhalo *B* by determining which later Dhalo contains the descendant of the most massive subhalo in *A* which survives at the next snapshot. In every case where a subhalo in *A* survives, we assign the descendant of the subhalo to Dhalo *B*. We repeat this process for all Dhaloes at each snapshot in decreasing order of redshift. This ensures that if any two subhaloes are in the same Dhalo at one snapshot, and both survive at the next snapshot, they will both be in the same Dhalo at the next snapshot.

This process produces a Dhalo catalogue for each snapshot. Each Dhalo contains one or more subhaloes and each subhalo may have a pointer to a descendant at some later snapshot. Any subhaloes in a Dhalo which survive at the next snapshot are guaranteed to belong to the same Dhalo at the next snapshot. This provides a simple way to identify a descendant for each Dhalo and defines the Dhalo merger trees. Fig. A3 shows an example of a Dhalo merger tree. The two smaller haloes *B* and *C* merge with a larger halo *A*. Halo *C* survives as a satellite for one snapshot before merging with the descendant of *A*. Halo *B* also becomes a satellite subhalo and then temporarily escapes from the parent halo before falling back in. At all times after the initial infall, it is considered to be part of the parent Dhalo.

This paper has been typeset from a TEX/LATEX file prepared by the author.