Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

Mon. Not. R. Astron. Soc. **427,** 2437–2449 (2012) doi:10.1111/j.1365-2966.2012.22111.x

Resolving subhaloes' lives with the Hierarchical Bound-Tracing algorithm

Jiaxin Han,[1,2,3]★ Y. P. Jing,[1] Huiyuan Wang[4,5] and Wenting Wang[1,2]

[1]*Key Laboratory for Research in Galaxies and Cosmology, Shanghai Astronomical Observatory, Shanghai 200030, China*
[2]*Graduate School of the Chinese Academy of Sciences, 19A Yuquan Road, Beijing 100049, China*
[3]*Institute of Computational Cosmology, Department of Physics, University of Durham, Science Laboratories, South Road, Durham DH1 3LE*
[4]*Key Laboratory for Research in Galaxies and Cosmology, University of Science and Technology of China, Hefei, Anhui 230026, China*
[5]*Department of Astronomy, University of Science and Technology of China, Hefei, Anhui 230026, China*

Accepted 2012 September 10. Received 2012 September 5; in original form 2012 August 27

ABSTRACT

We develop a new code, the Hierarchical Bound-Tracing (HBT for short) code, to find and trace dark matter subhaloes in simulations based on the merger hierarchy of dark matter haloes. Application of this code to a recent benchmark test of finding subhaloes demonstrates that HBT stands as one of the best codes to trace the evolutionary history of subhaloes. The success of the code lies in its careful treatment of the complex physical processes associated with the evolution of subhaloes and in its robust unbinding algorithm with an adaptive source-subhalo management. We keep a full record of the merger hierarchy of haloes and subhaloes, and allow the growth of satellite subhaloes through accretion from its 'satellite-of-satellites', hence allowing mergers among satellites. Local accretion of background mass is omitted, while rebinding of stripped mass is allowed. The justification of these treatments is provided by case studies of the lives of individual subhaloes and by the success in finding the complete subhalo catalogue. We compare our result to other popular subhalo finders and show that HBT is able to well resolve subhaloes in high-density environment and keeps strict physical track of subhaloes' merger history. This code is fully parallelized and freely available upon request to the authors.

Key words: methods: data analysis – methods: numerical – dark matter.

1 INTRODUCTION

In the hierarchical universe, cold dark matter haloes grow mainly through mergers with surrounding smaller haloes. After the merger, the imprints of progenitor haloes are not wiped out but in fact can survive for quite a long time as self-bound substructures called subhaloes (Ghigna et al. 1998; Moore et al. 1998; Klypin et al. 1999; Moore et al. 1999). Galaxies form inside dark matter haloes and can be traced by dark matter subhaloes after the merger. Because the non-linear growth of structure in the dark matter component can be well produced in *N*-body simulations, it has become a standard approach to build galaxy formation models on top of the dark matter halo merger history (see e.g. Baugh 2006; Benson 2010 for recent reviews). Constructing the full hierarchy of the merger history requires the identification and linking of dark matter subhaloes across cosmic time. Many algorithms have been developed to accomplish this job, all based on some of the following characteristics of a subhalo.

(i) It is an overdense region inside its host halo.
(ii) It is self-bound so that it is dynamically significant.
(iii) It was a halo before it merged into its current host halo.

Most subhalo finders utilize only contemporary particle distribution and focus on the first two characteristics, while the merger tree is constructed by a subsequent matching of subhaloes at different epochs. For example, the Hierarchical Friends-of-Friends (HFoF; Klypin et al. 1999) algorithm, which is an extension to the standard Friends-of-Friends (FoF; Davis et al. 1985) halo finder with multiple resolutions, makes use of the first characteristic only. AMIGA HALO FINDER (AHF; Knollmann & Knebe 2009), SUBFIND (Springel et al. 2001a) and SKID (Ghigna et al. 1998), which collect local overdense particles and then eliminate unbound particles, are based on the first two characteristics. Although these percolation-based algorithms are able to find subhaloes using a single simulation output, a strong resolution problem can arise in the central part of the host halo due to ambiguity of separating member particles of a subhalo from the background ones in the high-density region (Gill, Knebe & Gibson 2004; Muldrew, Pearce & Power 2011). In fact in a recent extensive halo-finder comparison project (Knebe et al. 2011), those participating finders based on configuration information only all struggle to recover substructures in the central high-density region of a host halo (see also Section 3.1). One way to improve the resolution of subhaloes in the high-density region is to use six-dimensional phase-space information, as done in e.g. the HIERARCHICAL STRUCTURE FINDER (HSF; Maciejewski et al. 2009) and ROCKSTAR (Behroozi, Wechsler & Wu 2011). Another way out

★E-mail: jxhan@shao.ac.cn

of this problem is to appeal to the third characteristic and utilize the evolutionary history of subhaloes. Because subhaloes are remnants of dark matter haloes, they can be identified by tracing the member particles of their progenitor haloes. The first attempt along this direction was Tormen (1997) who just considered the third characteristic. Later, in Tormen, Diaferio & Syer (1998), self-boundness was also added to define a subhalo. Only those haloes which fall directly into the final halo of interest were examined until the work of Giocoli et al. (2010, hereafter G10) where they extended their SURV code (Tormen, Moscardini & Yoshida 2004; Giocoli, Tormen & van den Bosch 2008) to include subhaloes inside subhaloes. The same method was also implemented in the 'MLAPM HALO TRACKER' (MHT; Gill et al. 2004).

As simple as the idea looks, the subhalo identification through tracing a merger history still faces several difficulties in practice. The most challenging aspect of the problem is how to trace subhaloes robustly over several orbital periods. Mass loss is the primary process associated with the subhalo evolution due to gravitational stripping and harassment. In such a simplified picture, once a particle becomes unbound to a subhalo, it no longer needs to be traced any more. However, as we will show, this kind of successive tracing is dangerous since an artificial loss of bound particles can accumulate through every tracing step, eventually triggering a runaway loss of a subhalo's bound particles. On the other hand, if one always traces all the particles from a subhalo's progenitor halo, a straightforward unbinding algorithm may also fail to find a self-bound structure once the subhalo has been substantially stripped compared to its progenitor halo, due to the large amount of unbound particles. Besides the stringent requirement on the robustness of tracing, second-order effects can lead to mass growth for satellite subhaloes. As we will show, both accretion and merger can continue to happen for satellites even within the virial radius of the host halo. Ignoring these processes will underestimate the mass of traced subhaloes.

In this work, we present a new tracing code (HIERARCHICAL BOUND-TRACING, HBT hereafter) to identify subhaloes and construct the merger tree. The key to our code is a robust unbinding algorithm together with an adaptive source-subhalo management. With these recipes we are able to walk through the cosmic age to capture every subhalo ever born, as will be described in Section 2. Our careful tracing of subhaloes' hierarchy enables us to apply the unbinding algorithm recursively, naturally allowing satellite accretion and merger. The success of HBT is demonstrated in Section 3 to have high completeness. Through comparison to a configuration space subhalo finder we also show that HBT subhaloes are in general more robust and more extended. The results are summarized in Section 4. HBT's implication for galaxy formation models and prospects are discussed in Section 5. Two concordant Lambda cold dark matter (ΛCDM) simulations, both with cosmological parameters $\Omega_{\rm M} = 0.3$, $\Omega_\Lambda = 0.7$ and $\sigma_8 = 0.9$, are used for the tests and comparisons in this work; one is a zoomed-in resimulation of a $10^{15}\,h^{-1}\,{\rm M}_\odot$ galaxy cluster with a particle mass $m_{\rm p} = 1 \times 10^8\,h^{-1}\,{\rm M}_\odot$ carried out with GADGET II (Springel 2005), and the other is a cosmological simulation with a boxsize $100\,h^{-1}$ Mpc using 512^3 particles (Jing & Suto 2002).

Our algorithm also makes use of characteristics (ii) and (iii) and it turns out that the first characteristic is automatically satisfied. The difference between our algorithm and that of G10 is mostly technical, mainly in the time direction of tracing a subhalo: HBT proceeds from the earliest epoch, naturally finding the full hierarchy of subhaloes together with extracting the merger tree in only one walk through the cosmic time, while G10 tries to figure out the subhalo hierarchy level-by-level by subsequently revisiting earlier snapshots. The MHT code depends on the 'MLAPM halo finder' (MHF; Gill et al. 2004) to first identify haloes as well as subhaloes from the halo formation time. Then the haloes and subhaloes are tracked in subsequent outputs. In HBT the growth, merger and stripping of haloes and subhaloes are handled in a unified way starting from the earliest resolved haloes, and the full hierarchy of the subhalo merger tree is resolved. HBT also stands out in the sense that it is the first time that various systematic issues in a tracing algorithm have been investigated and the tracing results have been carefully assessed.

2 ALGORITHM

2.1 Overall tracing algorithm

HBT starts the tracing of subhaloes from input halo catalogues for a sequence of simulation outputs.[1] In the current implementation we adopt the simple and widely used FoF halo finder to construct the input catalogues because it has been demonstrated that virialized haloes and subhaloes are included in these FoF haloes (Springel et al. 2001a), and also because it does not cut subhaloes near halo boundaries as for example a spherical-overdensity halo finder (Lacey & Cole 1994) would do. For our two ΛCDM simulations, the nominal linking length $b = 0.2$ is adopted. With halo catalogues in hand, the HBT algorithm can be summarized in one sentence: HBT builds and traverses the halo merger tree and finds the self-bound structure for every halo at every snapshot after its birth.

Specifically, the subhalo identification breaks into two steps: first selecting candidate particles which contain the members of the self-bound structure and then removing the unbound ones. We call the initial collection of the candidate particles, out of which the self-bound part gives rise to a subhalo, a *source-subhalo* (or source for short). In HBT we work with two types of subhaloes in each halo: a central subhalo which is the dominating subhalo within a host FoF halo, and satellite subhaloes which are the remaining ones if any. Starting from the highest redshift, haloes without progenitors are fed to an unbinding procedure as sources of their central subhaloes. At the next snapshot, the particles of these source subhaloes are tracked to identify their host haloes. If more than one source is found to reside in the same host, a halo merger is identified.[2] The progenitor sources are then unbound to select the most massive self-bound structure as the central subhalo, and others as satellites. The sources for the central subhaloes are updated to be the current host halo excluding particles from any self-bound satellites, and unbound again to allow for the growth of centrals. In this kind of tracing process, haloes without progenitors create new branches in the merger tree and give birth to new centrals, while mergers connect branches and transform centrals to satellites.

Since the particles of satellite subhaloes are traced from their progenitors, this is equivalent to assuming that no accretion of host halo's particles can happen for satellite subhaloes. In Appendix B, we explicitly test this assumption and show that local accretion has little effect on the tracing in the long term.

[1] Usually 60 snapshots for an ΛCDM simulation starting from an initial redshift where the first several haloes can be found, e.g. $z_{\rm ini} \sim 20$, is sufficient to achieve high completeness in the present-day subhalo catalogue. Interested readers can refer to Appendix C3 where we investigate how the algorithm depends on the time resolution of simulation outputs.

[2] It also happens that a source's particles are distributed into multiple host haloes. In this case, the source is split as described in Appendix C1.

We record the progenitor–descendant information as we proceed. This way we get the subhalo catalogues and the merger history of these subhaloes at the same time.

2.2 Tracing satellites robustly and efficiently

As we have mentioned before, it is the most challenging part for a tracing algorithm to trace satellite subhaloes robustly and efficiently. This is a problem mostly because unbinding can be fastidious about the quality of source subhaloes. Either too big or too small a source subhalo can lead to failure in unbinding. This can be understood as follows. The definition of boundness depends on the reference frame in which to calculate the kinetic energy and on the assembly of particles from which the potential energy is obtained. For a self-bound subhalo, the reference frame is defined to be at rest with respect to a certain 'centre' of the subhalo (e.g. centre of mass) which removes the bulk motion of the system, and the potential energy is summed over all the particles in the subhalo. Starting from a source subhalo with bound and unbound particles, one still seeks a centre which is roughly at rest with the final one, and uses this frame to calculate the binding energy. An overly large source subhalo with too many unbound particles would probably give a centre which deviates too much from the underlying subhalo, while an overly small one consisting of only a small portion of the real subhalo would give too shallow potential, both yielding severely biased result or even completely missing the subhalo. For example, as shown in Hayashi et al. (2003) a Navarro–Frenk–White (NFW) halo truncated to a radius of <0.77 times the scale radius is completely unbound. Thus, a source subhalo is expected to be a slightly enlarged assembly of particles containing the final subhalo. In addition, the unbinding procedure is required to be robust, especially if the source is strongly contaminated. We make efforts in both aspects to improve the robustness of HBT.

In HBT we implement a core-averaged unbinding algorithm designed to tolerate contamination. Starting from a source subhalo, unbound particles are removed iteratively till the bound mass converges. For each iteration, the reference frame is chosen to be the centre of mass and bulk velocity of an innermost core consisting of a certain fraction `CoreFrac` of the remaining particles with the lowest potential. This provides a robust estimate of the underlying true reference frame in the presence of contaminations. More detailed description of this algorithm can be found in Appendix A.

Besides improving unbinding, our HBT code updates satellite sources adaptively, to keep the amount of contaminating particles under control, while still maintaining a large enough reservoir of candidate particles. The idea is to reduce the size of the source conservatively every time a subhalo has been stripped to a fraction `CoreFrac0` of the current source mass $M_{\rm src}$. To ensure that the new source with mass $M_{\rm src2}$ is still larger than the current subhalo, we replace the source with the subhalo's progenitor found at the time it was first stripped to a fraction $\sqrt{\texttt{CoreFrac0}}$ of the original source mass. With the new source constructed this way, it is ready to see that the current subhalo automatically becomes the new source in the next loop of source updates, saving the trouble to search for new sources upon every update. The source for a subhalo before infall into a host halo is simply its host halo with all the satellite particles masked out. After infall this kind of source updates start to take effect all the way along the subhalo's history. Note that the removal of unbound particles not only reduces contamination, but also reduces the amount of calculation for unbinding. With source subhaloes updated this way, the unbinding procedure is always protected to work under the condition $\texttt{CoreFrac0} \times M_{\rm src} < M_{\rm sub} < \sqrt{\texttt{CoreFrac0}} \times M_{\rm src}$, substantially reducing the amount of work for unbinding and making the tracing more robust.

The `CoreFrac` parameter adopted by the unbinding procedure is initially `CoreFrac0`, and updated with $\texttt{CoreFrac} = M_{\rm src2}/M_{\rm src} \times \sqrt{\texttt{CoreFrac0}}$ every time we update the source. This is to ensure that the smallest subhalo out of $M_{\rm src}$ does not use an overlarge core when $M_{\rm src2} \ll M_{\rm src} \times \sqrt{\texttt{CoreFrac0}}$ due to discreteness in simulation output time. In the current implementation, we adopt $\texttt{CoreFrac0} = 0.25$.

In the development stage we also tried several other reference frames for unbinding, including the following.

(i) Centre of mass (CoM) frame. Take the centre of mass of source particles as centre and the average velocity as the bulk motion.

(ii) Centre of potential (CoP) frame. Take the potential-weighted centre of mass as centre and the potential-weighted average velocity as the bulk motion.

(iii) Minimum potential (MinPot) frame. Take the position of the particle with minimum potential energy as centre and average velocity of the source subhalo as the bulk motion. We avoid using the velocity of the minimum potential particle considering that it may have large dispersion with respect to the bulk motion.

For the source construction, we also tried two simple strategies. One is to always use the source from the infall without any updating (no-update), and the other is to update it aggressively by using the progenitor subhalo at snapshot $n-1$ as the source for a subhalo at snapshot n (immediate-update).

In Fig. 1 we compare the performance of various combinations of these unbinding and source construction recipes, by applying them to two subhaloes in the cluster simulation. Top panels show the case for tracing halo S51G86 (named after snapshot ID and halo ID just before the infall, with the merger mass ratio 0.0016 and the initial orbital circularity 0.16). The left-hand column is the 'no-update' regime, where the source is taken to be the progenitor halo at the infall, while the middle column does 'immediate update' and uses the subhalo at the previous tracing step as the source. It can be seen that the CoM frame method and the MinPot frame method have similar performance; both methods fail to identify more than a half of the bound population. The performance is worse at a smaller halocentric distance where the tidal force is stronger or when the bound part is smaller hence there are more contaminating particles. In general the core-averaged unbinding algorithm has the best performance, due to its ability to seek out a tight core to represent the majority of the bound particles. Still an overestimation of the core size (e.g. $\texttt{CoreFrac} = 0.5$ after $z = 0.2$ with no update) can cause an obvious drop in subsequent subhalo mass. As shown in the top left-hand panel, the subhalo gradually loses mass as it spirals into the centre of its host, reaches a minimum at pericentre passage near $z = 0.6$ and then regains mass as it moves out. Those regained particles are also from the initial infalling halo, and we found that they are also mainly bound particles before the pericentre passage. This *rebinding* of once unbound particles is suppressed in the top middle panel, where monotonic decrease in subhalo mass is forced by keeping only bound particles from the last snapshot as the source. In an extreme case as shown in the bottom panels for halo S43G11 (with merger mass ratio 0.02 and initial orbital circularity 0.10), the rebinding process can increase the subhalo's mass by a factor of 2.3 after pericentre, and suppression of the recapture can yield an underestimation of subhalo mass by a factor

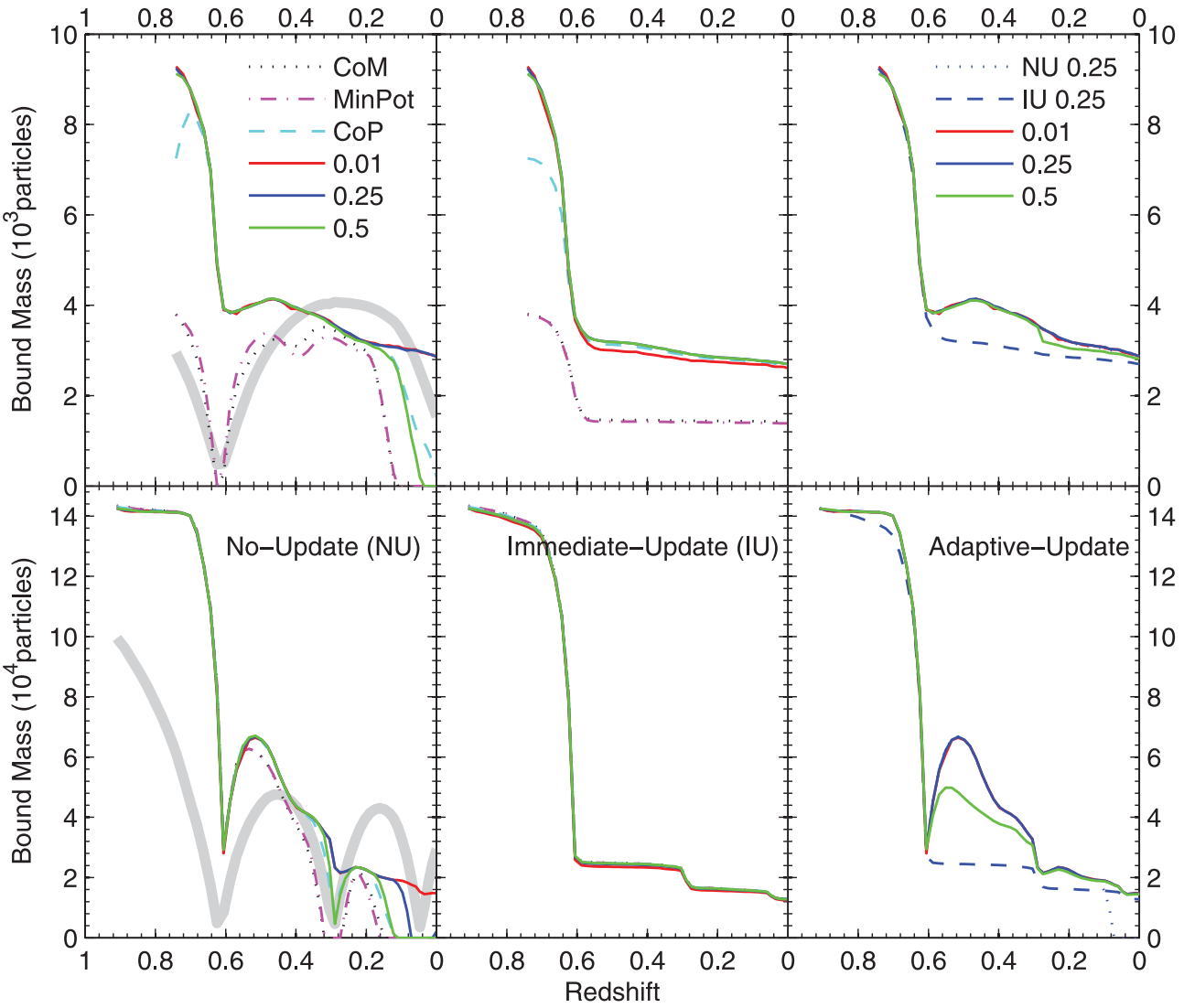

Figure 1. Robustness of various unbinding algorithms. Left: subhalo mass evolution calculated when the progenitor halo is used as a source without any updates during tracing. The solid lines are results of core-averaged unbinding, with different `CoreFrac` as given in the legend. The other three lines denote the results when three commonly used reference frames are adopted; the dotted line is for the centre of mass (CoM) frame, the dot–dashed line is for the MinPot frame and the dashed line is for the centre of potential (CoP) frame. These reference frames are defined in the text. The thick grey line shows the halocentric distance of the subhalo, in an arbitrary unit. Middle: the same as in the left-hand column, but the subhalo at the last tracing step is used as a source. Right: subhalo mass evolution calculated when a self-adaptive source is used. The solid lines are the results of self-adaptive core-averaged unbinding with different `CoreFrac0` parameters. Overplotted are the results of the core-averaged unbinding with no source update (the dotted line, overlapping with the solid blue line in the top panel; the same as in the left-hand column) and with immediate updates (the dashed line; the same as in the middle column), with `CoreFrac` = 0.25. Top panels are for halo S51G86, while bottom panels for halo S43G11.

of 2.8, comparing the left-hand and middle panels. The right-hand column shows the result when the adaptive update of source subhalo is applied. With the source size under control, the problem of an overly large core is avoided. Because $M_{\rm src}$ marks the maximum subhalo mass allowed, $M_{\rm src} < M_{\rm sub}/\texttt{CoreFrac0}$ shows that the maximum factor by which a satellite subhalo is allowed to grow through rebinding is $1/\texttt{CoreFrac0}$. Thus, a big `CoreFrac0` would still suppress mass growth, as in the case of the bottom right-hand panel with $\texttt{CoreFrac0} = 0.5$.

Fig. 2 further shows the reason for the performances of different unbinding algorithms. At the time of infall, using the reference frame of the final self-bound subhalo in the $\texttt{CoreFrac0} = 0.25$ unbinding result as the standard one, we check the difference of the initial estimation of the reference frames in each algorithm from that standard frame, and plot them normalized by the host halo virial radius $R_{\rm vir}$ and virial velocity $V_{\rm vir} = \sqrt{GM_{\rm vir}/R_{\rm vir}}$. The CoM, MinPot and CoP frames all have large velocity residuals with respect to the bound structure, although the MinPot frame has a negligible positional displacement.[3] However, the initial CoM frame also has a large positional displacement. The assignment of these displaced frames would inevitably result in the artificial loss of bound mass in these three algorithms.

The most bound particle of an HBT subhalo is almost always located at the centre of mass position of the core, with a displace-

[3] Note that we also adopted the average velocity as reference for the MinPot frame.

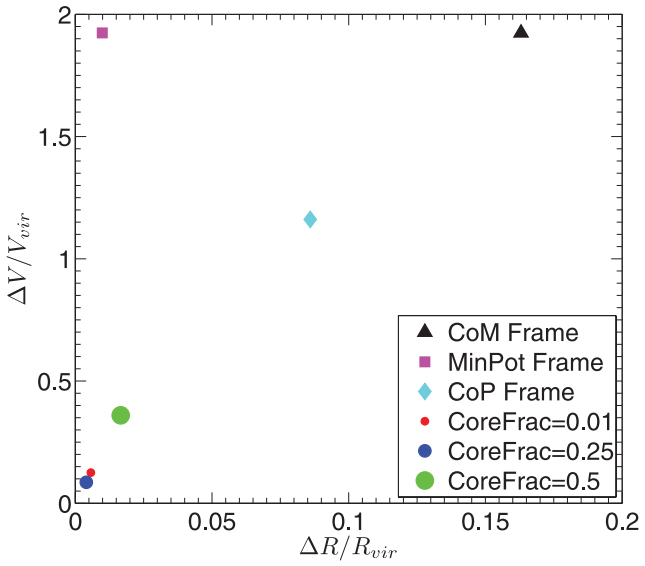

Figure 2. The reference frames in different unbinding algorithms for halo S51G86. For each algorithm, we plot the difference of the initial estimation of the origin and the relative velocity of the reference frame from the final reference frame found for the subhalo in the HBT code, normalized by the halo virial radius and virial velocity.

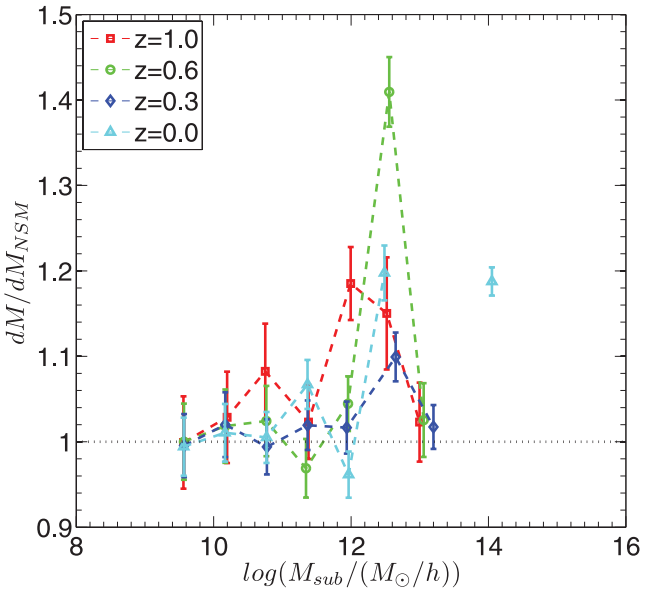

Figure 3. The effect of satellite merger. We plot the ratio of the total subhalo mass in each logarithmic mass bin between HBT results with (dM) and without (d$M_{\rm NSM}$) satellite merger for the resimulated cluster. The errors are propagated from Poisson errors in the total number of particles within each bin. Different colour lines with symbols are for different redshifts, while the horizontal dotted line is the 1:1 reference.

ment within one softening length, reflecting the definition of our boundness.

2.3 Allowing accretion and merger within satellite systems

Although the dominating physical process in satellite mass evolution is tidal mass loss, it has been shown in Simha et al. (2009) that mass accretion and merger is not terminated for satellite galaxies. Angulo et al. (2009) also find that the merger rate between satellite subhaloes and that between a satellite and a central subhalo are comparable for satellites smaller than 0.01 times the host halo mass, and that most of the satellite–satellite mergers happen between subhaloes which were once in central–satellite relation. These satellites, and 'satellite-of-satellites', as well as 'satellite-of-satellite-of-satellites', and so on, define hierarchical satellite systems which could persist for a long period up to many Gyr as observed in White, Cohn & Smit (2010), and these systems are places where satellite accretion and mergers can still happen.

In HBT we keep record of this 'sat-of-sat' hierarchy. Each central subhalo is aware of its satellites, and each satellite is aware of its sat-of-sats which it had accumulated before its infall and which are still alive. There are occasions when a subhalo is ejected from a host, which is quite common as shown in Lin, Jing & Lin (2003) (see also Gill, Knebe & Gibson 2005; Sales et al. 2007; Ludlow et al. 2009). In this case the ejected subhalo is also removed from the sat-of-sat list that contains it. Here we use the term 'sat-of-sat' instead of 'sub-in-sub' (see e.g. Springel et al. 2008) to emphasize that this relation between subhaloes is a historical or dynamical relation, which may not correspond to the spatial nesting relation in the latter case because of the separation of orbits due to host halo's tidal force or multiple satellite interactions. We allow the accretion of mass within each subhalo's sat-of-sats. This is achieved by implementing the unbinding procedure recursively: for a hierarchy of source subhaloes, starting from the highest level source subhalo, we feed the unbinding procedure with both particles from the current source subhalo and particles which are removed from recursive unbinding of lower level source subhaloes.[4] This means that the particles removed from a satellite can have a chance to be accreted into a higher level subhalo. When a satellite subhalo dies with the majority of its particles accreted by a higher level satellite subhalo, a satellite–satellite merger happens.

Ignoring the effect of satellite merger could result in the loss of subhaloes during tracing. For example, a subhalo with 7000 particles (or about 1/1000 of the host halo mass) is missing in the HBT result for the resimulation when satellite merger is switched off. This is triggered by a satellite major merger event between two satellites which were once in the central–satellite relation. After merger the particle kinetic energies are increased. If one still uses particles from only one source subhalo to do the unbinding, the potential would not be deep enough to bind the particles, resulting in the loss of the new subhalo.

In Fig. 3 we show the ratio of the total mass contained in each subhalo mass bin for the resimulated cluster, found by HBT with and without the satellite merger. This is equivalent to the ratio of the mass-weighted subhalo mass function $M_{\rm sub}{\rm d}N/{\rm d}\ln M_{\rm sub}$ (see e.g. Gao et al. 2004b for more details of the mass function). Although allowing merger among satellite subhaloes has almost no effect on small subhaloes, it can enhance the mass function at the high-mass end by 20 per cent, a fraction close to the fractional mass contribution from satellites to host haloes. The strong fluctuations at the high-mass end reflect the occurrence of significant satellite merger events.

[4] To avoid adding satellite particles multiple times, we define the source of a central subhalo to be the host halo excluding all the particles from its satellite *sources* when recursive unbinding is applied.

3 RESULTS AND COMPARISON

To show that HBT has a superior tracing ability to recover subhaloes in high-density regions and to compare with other subhalo finders, we have applied HBT to one test case provided by the 'Haloes Gone Mad' project (Knebe et al. 2011), to examine a subhalo's mass evolution. To show that HBT completely recovers the subhalo population, we compare the subhalo mass functions from HBT for our simulations with those from SUBFIND, as well as with a fitting formula. The comparison shows that HBT subhaloes are more massive than those from SUBFIND by 10–20 per cent, a fact that is also observed when comparing the size of subhaloes from the two codes. The reason for this difference is further revealed in the density profile, where SUBFIND subhaloes show sharp truncation near tidal radii while HBT subhaloes are more extended. We also show the one-to-one matching result between HBT and SUBFIND catalogues. HBT has also been compared with many other subhalo finders in the subhalo-finder comparison project (Onions et al. 2012), and is found to have good performance.

3.1 Subhalo mass stripping history

The test simulation we use from the 'Haloes Gone Mad' project is a head-on collision simulation of two NFW haloes with initial virial masses of 10^{12} and $10^{14}\,\mathrm{M}_{\odot}$, corresponding to 10^4 and 10^6 particles. The small halo is thrown right through the centre of the other halo. As shown in Knebe et al. (2011), all halo finders based on only configuration space information fail to find the subhalo at host centre, and even some finders based on position and velocity phase-space information give an unphysical result near the centre. For clarity, we choose two representative configuration-based finders (AHF and SUBFIND) and two phase-space-based finders (HSF and ROCKSTAR), and compare our HBT result in Fig. 4 with those found by the four finders. It can be seen that the HBT subhalo is complete from the start, robust near the halo centre, and clean as it moves out of the central region. As just stated, the two configuration-based finders fail to find the subhalo at the very central region of the host halo. HSF gives an unphysically high subhalo mass in the beginning (higher than the halo mass before the infall), which may be attributed to the inclusion of some local particles as discussed in Appendix B. AHF gives a much higher mass when the subhalo just passes the centre, which is not physical because, as can be seen from the right-hand panel, the maximum circular velocity $V_{\rm max}$ well exceeds that of the progenitor halo. A significant fluctuation in the $V_{\rm max}$ history is also observed for the subhalo found by ROCKSTAR near the host centre, indicating some ambiguity in capturing the bound mass by the finder. Furthermore, there is no obvious reason why the subhalo's mass should peak at 0.4 times the host virial radius as given by ROCKSTAR. In general HBT has comparable performance with HSF, while still has the advantage that all the HBT particles are strictly physically associated with the subhalo by design.

3.2 Subhalo mass function

In Fig. 5, we show the subhalo mass function found by HBT and SUBFIND for the resimulated cluster at $z = 0$. For reference we also plot the power-law fitting formula (Gao et al. 2004b; Springel et al. 2008; Angulo et al. 2009; G10)

$$\frac{\mathrm{d}N}{M_{\rm host}\mathrm{d}\ln(M_{\rm sub})} = N_0 M_{\rm sub}^{-0.9}, \quad (1)$$

where $M_{\rm sub}$ is the subhalo mass and $M_{\rm host}$ is the host halo virial mass. The normalization N_0 depends on the definition of the virial radius within which subhaloes are counted. For the virial relation predicted by spherical collapse model (see e.g. Bryan & Norman 1998 for a fitting formula), G10 found the normalization to be $N_0 = 10^{-3.03}(h^{-1}\,\mathrm{M}_{\odot})^{-0.1}$. We adopt the same virial definition throughout this paper, with the average density within the virial radius defined to be 101 times the critical density at $z = 0$, and will compare our subhalo mass function with the G10 result. It can be seen that the mass functions are consistent with the fitting formula

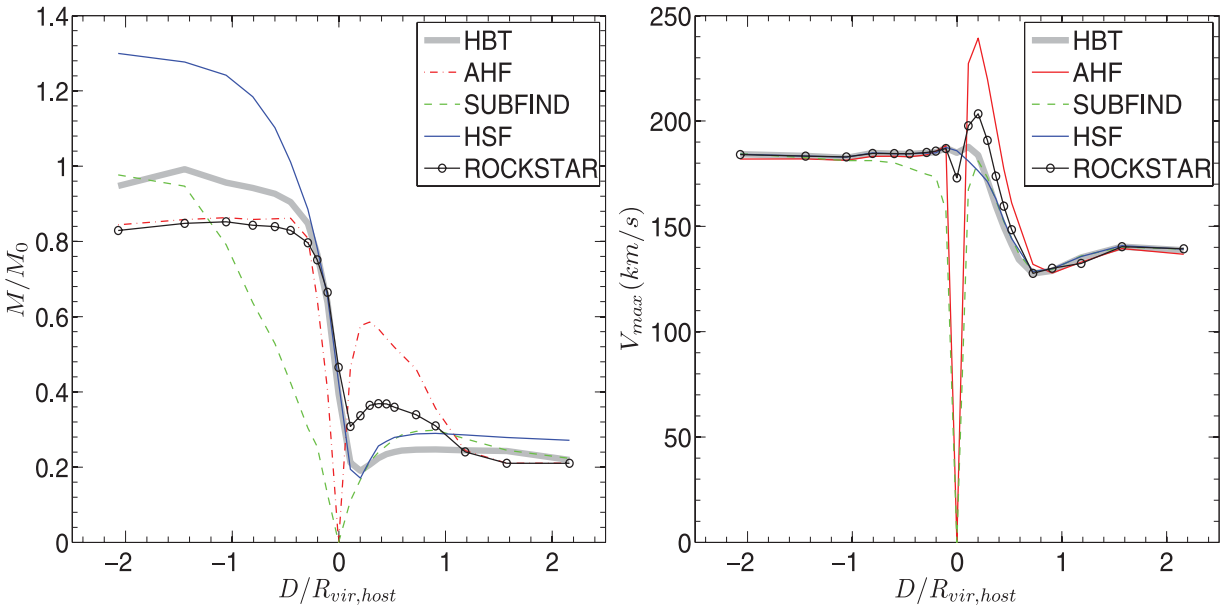

Figure 4. Mass stripping history of a subhalo falling through the centre of a host halo as found by HBT and the other four representative subhalo finders. Left: bound mass of the subhalo normalized by its virial mass before infall, as a function of its distance from host centre, normalized by the virial radius of the host halo. The subhalo moves towards the positive direction of the axis. Right: evolution of the maximum circular velocity of the subhalo. The circles on the ROCKSTAR line mark the time steps of the simulation outputs.

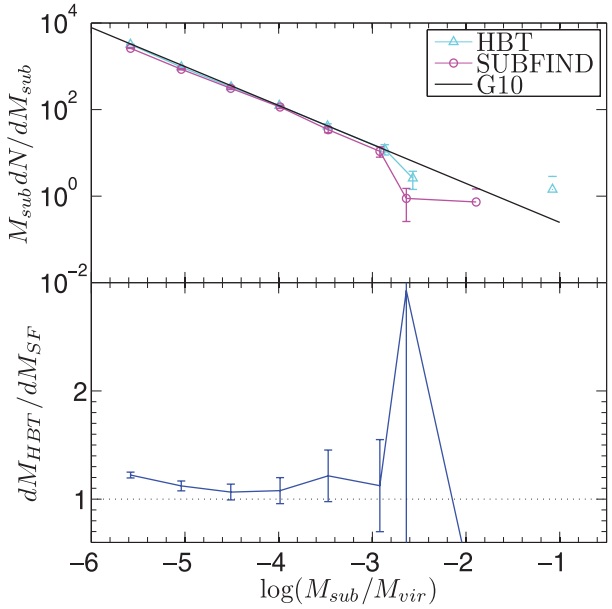

Figure 5. Subhalo mass function at $z = 0$ for the cluster resimulation. The top panel shows mass-weighted subhalo mass function as found by HBT and SUBFIND. The black solid line is the fitting formula of equation (3.1) with the normalization given by G10. The bottom panel plots the ratio of the total subhalo mass contained in each mass bin from the two codes, or equivalently the ratio of the mass-weighted subhalo mass functions. The horizontal dashed line is the 1:1 reference.

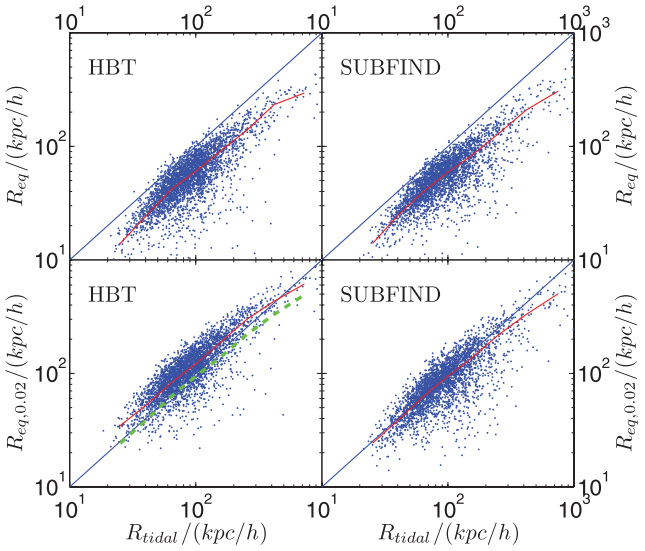

Figure 6. Size of subhaloes in the 100 most massive haloes in the cosmological simulation as found by HBT and SUBFIND. $R_{\rm eq}$ and $R_{\rm eq,0.02}$ are the radii where the subhalo density is equal to 1 and 0.02 times the background density. $R_{\rm tidal}$ is the subhalo tidal radius. We use only FoF haloes containing more than 1000 particles and subhaloes containing more than 100 particles. The red solid line in each panel marks the median value of $R_{\rm eq}$ (or $R_{\rm eq,0.02}$ according to the y-axis). In the lower left-hand panel, we overplot the median line of $R_{\rm eq,0.02}$ of SUBFIND result with a green dashed line.

from G10, with HBT's subhaloes being more massive than the result of SUBFIND by 10–20 per cent.

3.3 Size of subhaloes

The size of a subhalo is trimmed by the tidal force from the host halo, and can be estimated by a tidal radius which is defined as the radius of the satellite subhalo at which its self-gravity equals the tidal force of the host halo. In the limit $R_{\rm tidal} \ll D$ where D is the halocentric distance of the subhalo, the tidal radius is estimated as (Binney & Tremaine 1987; Tormen et al. 1998)

$$R_{\rm tidal} = D \times \left[\frac{M_{\rm sub}}{\left(2 - \frac{{\rm d}\ln M_{\rm host(<D)}}{{\rm d}\ln D}\right) M_{\rm host(<D)}} \right]^{1/3}. \quad (2)$$

For singular isothermal density profile, it is easy to check from equation (3.2) that the local density of the host halo is equal to that of the satellite at tidal radius. Thus, a radius of equality $R_{\rm eq}$ can be defined at which the density of the satellite is the same as the local density of the host. For realistic density profiles, the tidal radius may differ from $R_{\rm eq}$ by a factor of order unity. In Springel et al. (2008) they found that the subhalo tidal radius is equal to a radius $R_{\rm eq,0.02}$ where the subhalo's density falls below 0.02 times the local host density. We compare the tidal radius to $R_{\rm eq}$ and $R_{\rm eq,0.02}$ for both HBT and SUBFIND results in Fig. 6, using the 100 most massive FoF haloes in the cosmological simulation. To avoid resolution effect, we use only FoF groups with more than 1000 particles and subhaloes with more than 100 particles. It can be seen that although the $R_{\rm eq,0.02}$ is very close to $R_{\rm tidal}$ for SUBFIND result, it is in general 20–30 per cent larger in the HBT case. Moreover, for $R_{\rm eq}$ the two codes give consistent results. This reflects a puff-up in the outer region of subhaloes in the HBT result, or rather a sharp cut-off in the SUBFIND result, due to particle division using density saddle points. We show this explicitly in the next subsection.

3.4 Density profile

There is almost no difference in the density profile between central subhaloes found by HBT and SUBFIND. Moreover, for satellite subhaloes, especially when they reside in the central region of host haloes, HBT gives a much more extended profile. Fig. 7 compares the density profiles of the most massive satellite of the cluster found with the two codes. This subhalo is resolved to have 10^6 particles or 1/10 the virial mass of the cluster. It is located at $0.3R_{\rm vir}$ from the cluster centre. As expected, a sharp cut is observed for the SUBFIND profile near its tidal radius, while HBT gives a subhalo mass which is 6.5 times large. This is due to HBT's prior knowledge of source particles beyond the tidal radius while SUBFIND can only identify those particles within tidal radius, as already seen in Section 3.3.

3.5 Cross-match between HBT and SUBFIND

To see if the two codes find the same set of subhaloes, we try to match the members in the two catalogues. Given one target subhalo in hand, we search host subhaloes for its member particles in the other catalogue. For each host subhalo, we calculate a boundness-weighted summation of the number of matched particles, with the most bound particle in the target subhalo having the largest weight. The host subhalo with the biggest summation is selected as the target's correspondence in the other catalogue. We do the match from HBT to SUBFIND and vice versa. We classify the subhaloes into three categories according to the two matches: bilaterally matched subhalo when the subhalo and its correspondence are matched to each other; unilateral subhalo when the correspondence cannot be

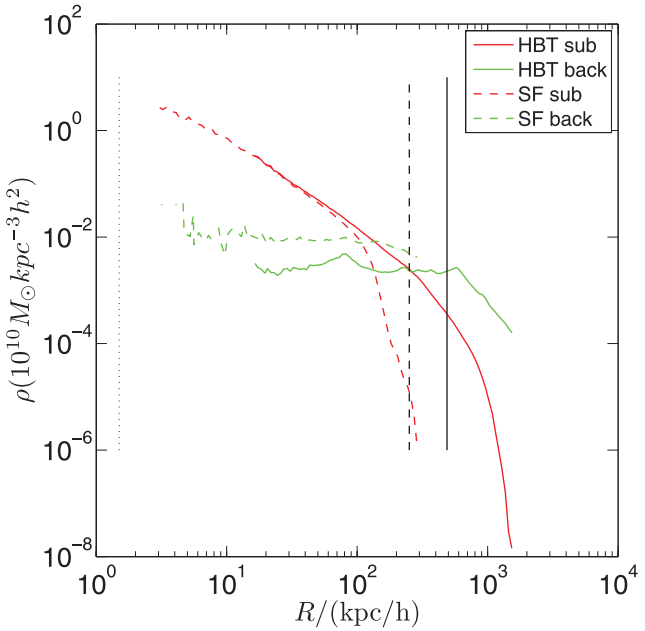

Figure 7. Density profiles for the most massive satellite subhalo of the resimulated cluster as found by HBT and SUBFIND. Red lines are for particles that are bound to the subhalo and green lines are the total density of all the other particles. The vertical black lines mark the tidal radius of the subhalo. Solid lines are the HBT result, while dashed ones are the SUBFIND result. The vertical dotted line on the left marks the smoothing length of the simulation.

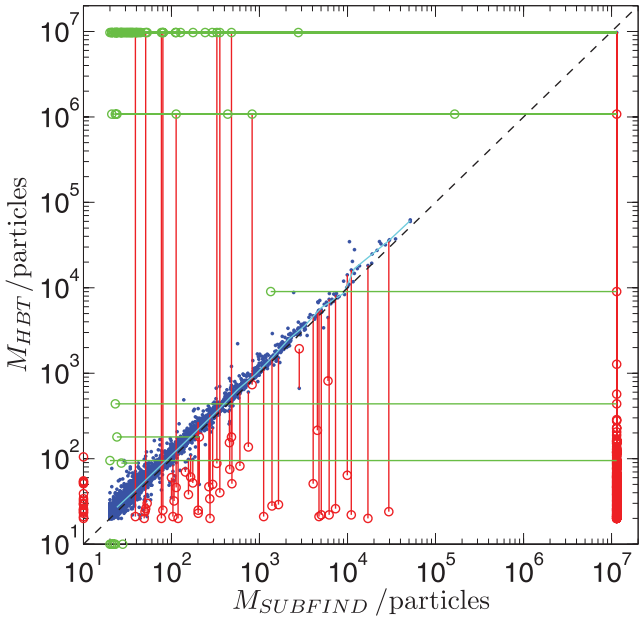

Figure 8. Cross-match between HBT and SUBFIND subhaloes for the cluster. Each subhalo is plotted according to its mass and the mass of its correspondence. Blue dots are bilaterally matched subhaloes (see the text for definition). The solid cyan line gives the median masses of the bilateral subhaloes. Red circles are unilateral HBT subhaloes and green squares are unilateral SUBFIND subhaloes, except for those on the axes which are unmatched ones. For each unilateral match, we also draw a line connecting that subhalo to its correspondence. The point at the top right-hand corner is the central subhalo.

matched back to the subhalo; and unmatched subhalo when no correspondence can be found in the other catalogue.

The match result for the galaxy cluster is shown in Fig. 8. The majority of subhaloes are bilaterally matched, accounting for 82 per cent HBT subhaloes and 97 per cent SUBFIND subhaloes. 90 per cent of the bilateral SUBFIND subhaloes have more than 95 per cent of their particles shared with their HBT correspondences. The median line points out 10–20 per cent increase in mass for HBT subhaloes compared to their SUBFIND correspondences. Only less than 1 per cent subhaloes belong to the unmatched category, with their masses being lower than 100 particles.

Unilateral subhaloes are discussed extensively in Appendix D. Here we only focus on one particularly interesting case. When two haloes with comparable masses merge, it is likely that the resulting halo contains two subhaloes with comparable peak densities, making it difficult to distinguish which is superior and which is subordinate without appealing to progenitor information. Especially when these two subhaloes are close to each other forming a high-density environment, the difficulty of resolving and weighing them are elevated. We give an example of such a binary system in Fig. 9 from the cosmological simulation. Two satellite subhaloes sit at the centre of the image resulting from a 1:2 merger of two progenitor haloes. This binary satellite system is found near the boundary of the host halo and thus closely resembles an individual halo. Most of the surrounding particles are bound to both of the density peaks. SUBFIND takes the peak from the smaller halo which has higher central density and associates the surrounding particles with it to make a parent subhalo, leaving the core from the bigger halo as its sub-in-sub, while HBT gives a reversed sub-in-sub hierarchy which is consistent with the merger hierarchy. When we match these two subhaloes from SUBFIND to HBT, obviously the small subhalo in SUBFIND is matched to the big one in HBT. And because the surrounding mass outweighs the core in the big subhalo of SUBFIND, this one is also matched to the big subhalo in HBT. For the same reason, the two HBT subhaloes are also matched to the big SUBFIND subhalo. This may be of particular importance when constructing merger histories. A false switch between two branches of a tree would happen as a result of the switch between the sub-in-sub relation of the binary. With the help of subhaloes' merger history, HBT is able to give more physical weights to the dominance of the two cores and avoid this kind of faulty links in the merger tree. In Fig. 8, these binary subhaloes would produce two adjacent perpendicular lines connecting near the 1:1 line if the surrounding particles could outweigh one core, or a single line if one subhalo is missing. A dedicated study of how major-mergers challenge different subhalo finders can be found in Behroozi et al. (in preparation).

4 SUMMARY

In this work, we have managed to develop a tracing algorithm to produce a complete and physically motivated subhalo catalogue. In our HBT algorithm, subhaloes are divided into two types where central subhaloes grow via accretion from the host halo and from satellites inside the same host while satellite subhaloes can only accrete from within their satellite-of-satellites. We keep a full record of subhaloes' merger hierarchy and apply the unbinding algorithm hierarchically to enable satellite accretion and merger. While omitting satellite accretion can result in 20 per cent loss in mass for massive satellites, individual tests of subhalo mass-loss history and statistical comparison with SUBFIND which is based on local density search show that local accretion of background particles have only temporary effect and is negligible for satellite subhaloes.

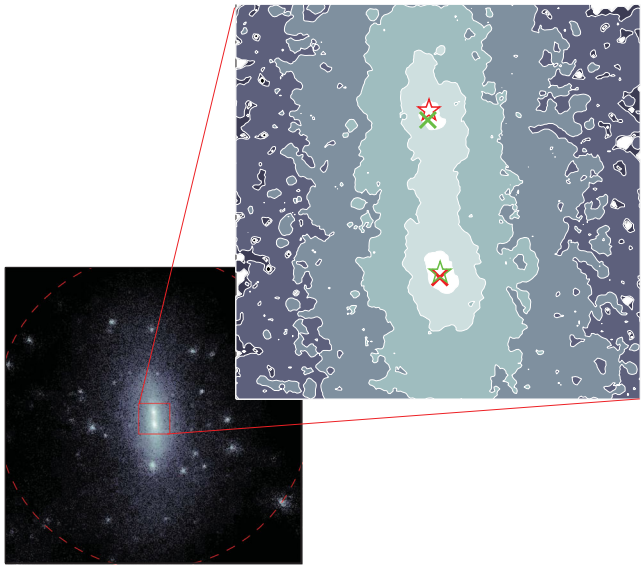

Figure 9. An example of sub-in-sub switch. These are two subhaloes forming a binary system near the boundary of the host halo. The lower left-hand panel shows the projected squared density of this system. The red dashed circle in the lower left-hand panel marks the virial radius of the system. The upper right-hand panel shows the contour map of the density field for the very inner region. Red symbols mark the most bound particles of two HBT subhaloes; green symbols mark those of SUBFIND subhaloes. Pentagrams represent the superior subhaloes and crosses stand for subordinate subhaloes. The masses for the upper and lower subhaloes in the image are 94 468 (488) and 17 541 (111 884) particles in the HBT (SUBFIND) case.

Because of the large dynamic range in the evolution of subhaloes' masses, a robust unbinding algorithm is required to trace subhaloes long enough. We have proposed core-averaged unbinding algorithm together with self-adaptive update of source subhaloes to accomplish this job. The core-averaged unbinding algorithm succeeds in its ability to give quick and accurate estimation of the centre and bulk velocity of the subhalo, allowing safe removal of unbound particles. Self-adaptive update of source subhaloes keeps the dynamic range in a safe region for unbinding while allowing rebinding of stripped particles. We find that successive tracing of satellites without allowing rebinding can cause up to a factor of 3 suppression in the subhalo mass at apocentre passage. Application of HBT to a contrived simulation also demonstrates its superb performance in robustly recovering the subhalo's stripping history.

Since our algorithm utilizes historically constructed source subhaloes rather than locally collected sources using density and position thresholds, we can resolve satellite subhaloes well even in the central high-density region of haloes while SUBFIND may have severe spatial truncation for subhaloes in halo centre. Even for small satellites our subhalo mass function is about 15 per cent higher than SUBFIND. The size of subhaloes found by HBT extends 20–30 per cent larger than SUBFIND characterized by $R_{\rm eq,0.02}$.

Because we do not need to do density interpolation or spatial searching to construct source subhaloes, HBT runs fast. The HBT code is written in C and has been fully OpenMP parallelized both for cosmological simulations where the parallelization is on halo level and for high-resolution resimulations where the parallelization is done on particle level; a hybrid MPI/OpenMP version is under development and will be available in the near future. It would also be straightforward and efficient to integrate HBT to a simulation code for on-the-fly high-resolution subhalo finding and merger tree output. In that case, HBT would also not be limited by the number of snapshots in the simulation output.

5 DISCUSSION

Much of the treatment in HBT is more physical than mathematical. Its complexity lies in the physical process involved and its success reflects its correct understanding of the subhalo's evolution. This code is obviously not applicable to hot dark matter simulations, where structure formation is top–down rather than bottom–up. We would also expect HBT to have some difficulty when applied to warm dark matter simulations, where fragmentation is a vital process in forming structures (Bode, Ostriker & Turok 2001), though our splitting algorithm would alleviate the problem. In fact, HBT can be regarded as a low-level semi-analytic model for cold dark matter subhaloes. Its high resolution and physical particle partitioning guarantees that the merger trees built by HBT have much fewer lost nodes or false links. This makes HBT an ideal choice for building merger trees for semi-analytic models of galaxy formation. Besides, the tracking nature makes HBT easily extensible to find not only bound subhaloes, but also more general structures such as streams. A combination of HBT's tracking ability with the STructure Finder's (STF; Elahi, Thacker & Widrow 2011) stream identification algorithm is under development to trace streams (Elahi et al., in preparation).

One example revealing HBT's advantage for galaxy formation models is the 'sub-in-sub switch' case shown in Fig. 9 and discussed in Section 3.5. When two subhaloes of comparable mass form close pairs, HBT is superior in its ability to partition the surrounding particles according to their origins. This is important because the surrounding particles, while being cobound by the binary subhalo and can be assigned arbitrarily to either member, play a vital role in establishing links to subhaloes' progenitors and descendants, especially when they outweigh the cores of the binary. HBT's way of partition would guarantee that these surrounding mass obey the progenitor–descendant relation, while a partition without knowing the origin of these particles can easily lead to an incorrect link of the cores to the progenitor haloes.

One future direction for the improvement of HBT code, and other subhalo finders as well, would be to find a better definition for subhaloes. Even with our ability to clearly construct source subhaloes near halo centre, subhaloes can still appear overstripped near pericentre after which some stripped mass get rebound, posing a remaining 'resolution' problem. We argue this remaining 'resolution' problem to be due to the operational definition of subhaloes as instantaneously self-bound structures, which is currently adopted by almost every subhalo finder. This definition only guarantees coherence for a structure in isolation and in the collisionless limit of its member particles. Putting aside the structural evolution of the subhalo itself, there is still an evolving background which lowers the gravitational potential of the subhalo while at the same time produces tidal force. In Shaw et al. (2007) both the tidal force from the host and the gravitational potential of stripped particles are considered to improve the coherency of subhalo definition. They find that the contribution from the tidal force and that from the lowered potential due to background particles roughly offset each other, with a slight increase for the mass of subhaloes in the inner region of a halo.

It should be noted that tidal force is not always disruptive. This could easily be illustrated in the following simple picture. Consider two test particles with infinitesimal mass moving on the same elliptical orbit around a central mass M, starting from apocentre

with a small time delay dt between them. The relative velocity is then $\mathrm{d}\boldsymbol{v} = (GM/r^3)\boldsymbol{r}\mathrm{d}t$. Hence the internal kinetic energy of the system varies as $\mathrm{d}v^2 \propto r^{-4}$, reaching a minimum at apocentre and maximum at pericentre. Because the internal gravity of the two test particles can be ignored, the change in the internal kinetic energy is solely from work done by tidal force. It is ready to see that the tidal work helps to reduce the relative kinetic energy of the two particles from pericentre to apocentre. As a result, particles previously marked as unbound to a subhalo can become bound again as the subhalo moves to the outer region of its host. This has been seen in our case study of the evolution of individual subhaloes in Section 2.2, where strong rebinding is observed as satellites move out from the centres of haloes.

We comment that in Shaw et al. (2007) even though tidal energy has been included in the unbinding procedure, they remove velocity outliers before adding tidal energy, still failing to allow these high-velocity particles to be recaptured through tidal work.

ACKNOWLEDGMENTS

We thank Volker Springel for making the SUBFIND code available, and the anonymous referees for lots of helpful comments. This work in Shanghai is supported by NSFC (10821302, 10878001, 11033006), by 973 Programme (No. 2007CB815402) and by the CAS/SAFEA International Partnership Programme for Creative Research Teams (KJCX2-YW-T23). HYW is supported by NSFC 11073017. During the final stage of this work, JXH is supported by the European Commission's Framework Programme 7, through the Marie Curie Initial Training Network Cosmo-Comp (PITN-GA-2009-238356).

REFERENCES

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983
Barnes J., Hut P., 1986, Nat, 324, 446
Barnes J. E., Hut P., 1989, ApJS, 70, 389
Baugh C. M., 2006, Rep. Prog. Phys., 69, 3101
Behroozi P. S., Wechsler R. H., Wu H.-Y., 2011, ApJ, submitted (arXiv:1110.4372)
Benson A. J., 2010, Phys. Rep., 495, 33
Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton Univ. Press, Princeton
Bode P., Ostriker J. P., Turok N., 2001, ApJ, 556, 93
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Elahi P. J., Thacker R. J., Widrow L. M., 2011, MNRAS, 418, 320
Gao L., White S. D. M., Jenkins A., Stoehr F., Springel V., 2004b, MNRAS, 355, 819
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Gill S. P. D., Knebe A., Gibson B. K., 2004, MNRAS, 351, 399
Gill S. P. D., Knebe A., Gibson B. K., 2005, MNRAS, 356, 1327
Giocoli C., Tormen G., van den Bosch F. C., 2008, MNRAS, 386, 2135
Giocoli C., Tormen G., Sheth R. K., van den Bosch F. C., 2010, MNRAS, 404, 502 (G10)
Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, ApJ, 584, 541
Jing Y. P., Suto Y., 2002, ApJ, 574, 538
Klypin A., Gottlöber S., Kravtsov A. V., Khokhlov A. M., 1999, ApJ, 516, 530
Knebe A. et al., 2011, MNRAS, 415, 2293
Knollmann S. R., Knebe A., 2009, ApJS, 182, 608
Lacey C., Cole S., 1994, MNRAS, 271, 676
Lin W. P., Jing Y. P., Lin L., 2003, MNRAS, 344, 1327
Ludlow A. D., Navarro J. F., Springel V., Jenkins A., Frenk C. S., Helmi A., 2009, ApJ, 692, 931
Maciejewski M., Colombi S., Springel V., Alard C., Bouchet F. R., 2009, MNRAS, 396, 1329
Moore B., Governato F., Quinn T., Stadel J., Lake G., 1998, ApJ, 499, L5
Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19
Muldrew S. I., Pearce F. R., Power C., 2011, MNRAS, 410, 2617
Onions J. et al., 2012, MNRAS, 423, 1200
Sales L. V., Navarro J. F., Abadi M. G., Steinmetz M., 2007, MNRAS, 379, 1475
Shaw L. D., Weller J., Ostriker J. P., Bode P., 2007, ApJ, 659, 1082
Simha V., Weinberg D. H., Davé R., Gnedin O. Y., Katz N., Kereš D., 2009, MNRAS, 399, 650
Springel V., 2005, MNRAS, 364, 1105
Springel V., White S. D. M., Tormen G., Kauffmann G., 2001a, MNRAS, 328, 726
Springel V., Yoshida N., White S. D. M., 2001b, New Astron., 6, 79
Springel V. et al., 2008, MNRAS, 391, 1685
Tormen G., 1997, MNRAS, 290, 411
Tormen G., Diaferio A., Syer D., 1998, MNRAS, 299, 728
Tormen G., Moscardini L., Yoshida N., 2004, MNRAS, 350, 1397
White M., Cohn J. D., Smit R., 2010, MNRAS, 408, 1818

APPENDIX A: UNBINDING ALGORITHM

The core-averaged unbinding in HBT is done in the following steps.

(i) Starting from an initial assemble of N particles of a source subhalo.

(ii) We calculate the gravitational potential ψ for each particle from the contribution of all the other $N - 1$ particles. We use the Barnes–Hut tree algorithm (Barnes & Hut 1986, 1989) as implemented in GADGET (Springel, Yoshida & White 2001b; Springel 2005) to calculate the potential. We then find a fraction (set by a parameter `CoreFrac`) of particles with the lowest potential, and call it as the lowest potential core.

(iii) We calculate the kinetic energy K of each particle, including Hubble flow, with respect to the average velocity and centre of mass of the lowest potential core. Then we remove any particles with a positive total energy $K + \psi > 0$.

(iv) If the remaining number of particles N_b is below the user-desired mass limit `NBoundMin`, the unbinding procedure stops with no subhalo found. Otherwise, if N_b agrees with N within some accuracy `MassPrecision`, unbinding stops with these N_b particles as a subhalo. In the other case, take the remaining N_b particles as an initial assemble of particles and repeat steps (ii)–(iv).

The parameter `MassPrecision` is introduced to improve the efficiency of unbinding, which is the most time-consuming part of the entire code. Strictly speaking the definition of self-boundness requires the iterative removal of unbound particles to be 100 per cent complete, i.e. the iteration can only stop when the number of bound particles N_bound equals the number of remaining particles N. However, as we have discussed in Section 5, self-boundness is not a perfect definition to identify particles associated with a subhalo. Practically, it would be enough to just give an estimation of the bound structure which is fairly close to the final self-bound part.

We investigated the speed of convergence for several randomly selected subhaloes and find that the bound mass converges quickly with the number of iterations towards the final self-bound mass, as shown in Fig. A1. After several iterations, the majority of unbound particles have already been removed. Further iterations only refine the bound mass to a higher accuracy, with smaller and smaller change in mass as the iteration goes on. An estimation of the

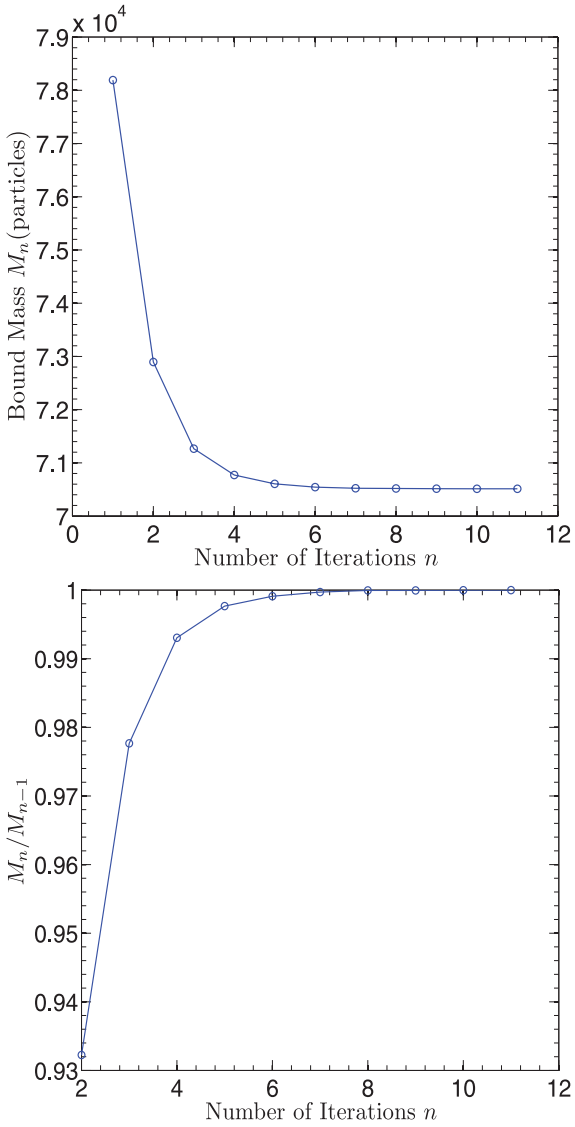

Figure A1. Typical convergence curve of a subhalo during unbinding. Top: the remaining bound mass after each iteration; Bottom: estimated precision of the bound mass after each iteration.

precision of bound mass can be obtained as the ratio of bound mass at two subsequent iterations. By stopping the iteration at some precision threshold `MassPrecision`, a lot of iteration steps can be saved while keeping the mass estimation accurate to the preset accuracy. We adopt `MassPrecision` = 0.995 in the current implementation of HBT.

APPENDIX B: POSSIBLE LOCAL ACCRETION

The physical assumption under tracing algorithms is that the mass accretion of satellite subhaloes can be ignored within a host halo, so that one needs only to follow the remnants of infalled haloes. In Fig. B1, we give a direct test of this assumption. After extracting a self-bound subhalo from its adaptive source, we search for and add any additional bound particles located within 2 times its virial radius (defined as if the subhalo is a halo). Except in the first several snapshots (and the first apocentre passage in the S43G11 case) when the subhalo is around the boundary of its host, no accretion of local background mass is found.

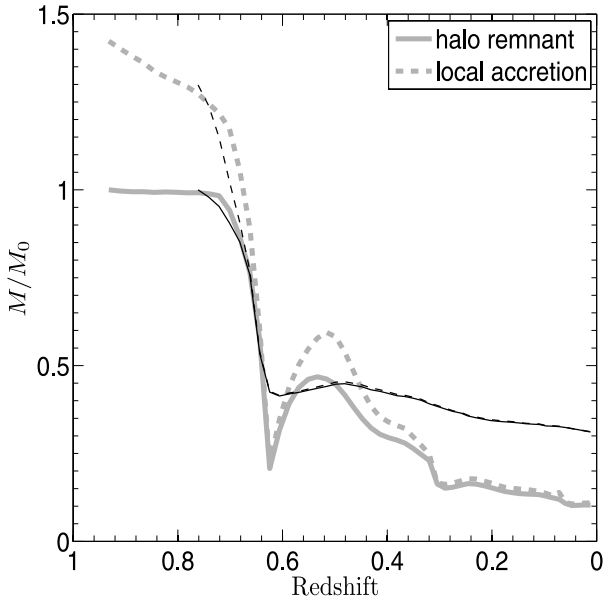

Figure B1. Effect of local background particles versus infalled source subhalo. The solid line shows the bound mass from the infalled halo, as done in the right-hand panel of Fig. 1 with `CoreFrac0` = 0.25. The dashed line shows the mass after adding bound local particles. The masses are normalized by the subhalo's mass at infall. Thick grey and thin black lines are for haloes S43G11 and S51G86, respectively.

APPENDIX C: MAXIMIZING THE RESOLUTION OF HALOES AND SUBHALOES

C1 Splitting algorithm

There are several conditions under which one may need to split a source subhalo into parts. One case is that a source subhalo may contain additional substructures which are unresolved given the time and spatial resolution of the simulation outputs, or from a casual link of two haloes as one. These unresolved substructures may separate from the hosting source subhalo later in orbits, due to orbital parameter difference or tidal forces of nearby structure. Moreover, a strong tidal force can break up a nearby smooth halo. In these cases, a source subhalo may have multiple descendants hosted by several haloes. We implement the following splitting algorithm to take this into account.

For each source subhalo, we partition its particles into fragments according to their current host haloes. Fragments with particle numbers smaller than `NSrcMin` are discarded. Here, `NSrcMin` is a parameter controlling the mass resolution of source subhaloes. We choose this to be equal to `NBoundMin` in our implementation. If all the fragments are discarded, discard this source subhalo. The subhalo out of the most massive fragment is taken as the 'main descendant' of this source subhalo. Other remaining fragments are called splinter source subhaloes, the subhaloes directly out of which are named splinter subhaloes. We keep a record of both the main descendant and the splinter information, yielding a merger tree which can have 'downward branches'. Note that because we do not appeal to spatial clustering within haloes to identify subhaloes, our tracing algorithm can only resolve splitting when it happens spanning over several haloes. Within one halo even if a subhalo breaks into two obvious part, as long as they are still bound as a whole we have no way to split them in HBT.

The mass function of the biggest splinters (excluding the main descendant) in each splitting event decays approximately as

$\mathrm{d}N/\mathrm{dln}\,m \propto m^{-2.5}$, much steeper than the slope of subhalo mass function $\mathrm{d}N/\mathrm{dln}\,m \propto m^{-0.9}$. It is ready to see that these splinters are primarily small subhaloes near the resolution limit.

C2 Quasi-halo treatment

We take those subhaloes for which we cannot find their host FoF haloes as being in a 'background halo' when assigning hosts in tree constructing algorithm as well as when splitting source subhaloes. This way we can find subhaloes which do not have a known host halo. Their source subhaloes can be regarded as 'quasi-haloes' which are not identified by the halo finder, but they still have concentrated structure and progenitor–descendant link between neighbouring snapshots. These quasi-haloes are haloes fluctuating around the resolution limit. The biggest quasi-haloes identified have about 50 particles when haloes are filtered by a mass limit of 10 particles in our implementation. Although they could be split or ejected subhaloes, those who can grow to a meaningful size are almost all haloes at their earliest stage of growth, i.e. haloes which have just become resolvable. The introduction of quasi-haloes is intended to recover the lost nodes in the merger tree due to halo-finder pitfalls, but has little effect on the tree even if omitted because they come as primarily early fluctuations.

C3 Time resolution dependence

In our tracing algorithm, one problem of concern is whether the time resolution of simulation outputs would affect the tracing result. The effect of time resolution is twofold: time resolution translates into the resolution of halo growth history; it also changes the dynamic range for unbinding thus affecting its robustness.

First, there could be unresolved satellites as well as an unresolved growth of satellites. Obviously those haloes born between two subsequent outputs and immediately becoming satellite subhaloes in the second snapshot are not recorded in halo catalogues and their descendants as independent subhaloes are missed. However, since these haloes get stripped right after they become discernible, they do not have time to grow and are always fairly small haloes. Moreover, the better the time resolution and the smaller the halo mass limit, the smaller are these unresolved haloes. For those haloes which merge between two subsequent snapshots, the growth of satellite haloes after the first snapshot and before merger are not captured either. Thus, one would expect a decrease in subhalo population with decreasing time resolution.

To quantify this effect, we compare the subhalo mass function $M_{\rm host}^{-1}\mathrm{d}N/\mathrm{d}\ln M_{\rm sub}$ identified with different time resolutions. Our cosmological simulation has 60 outputs from $z = 15.02$ to 0 equally spaced in log space of the scale factor. We dilute these outputs by keeping one snapshot after skipping every a few number of snapshots. Subhaloes are identified using these diluted snapshots with HBT. It is found that the diluted subhalo mass function has the same shape as the best resolved one, with the ratio of them well fitted by a constant and depending weakly on host halo mass, as long as the time step used for tracing is not too large. In Fig. C1, we examine the relative amplitude of the diluted subhalo mass function to the best resolved one (with $\Delta\ln a = 0.0468$) at five different redshifts for the cosmological simulation. The relative amplitude A as a function of time resolution can be well fitted by a Gaussian function

$$A = A_0 \mathrm{e}^{-[(\Delta\ln a)/b]^2}, \qquad \text{(C1)}$$

with parameters A_0 and b, where $\Delta\ln a$ is the difference in $\ln a$ between subsequent snapshots with a being the scale factor. The parameter A_0 represents the relative amplitudes extrapolated to infinitesimal time resolution. The fitted A_0 are all fairly close to unity, increasing from 1.001 ± 0.001 to 1.003 ± 0.001 from $z = 0$ to 2.1, indicating that the best resolved mass function is over 99.7−99.9 per cent complete. The parameter b depends on the scale factor as $b = 1.29a + 0.51$. Equation (C1) provides a handy reference in assessing the completeness A/A_0 for subhalo populations from HBT, although the exact completeness for a particular application would also depend on the mass resolution of the halo and subhalo catalogues.

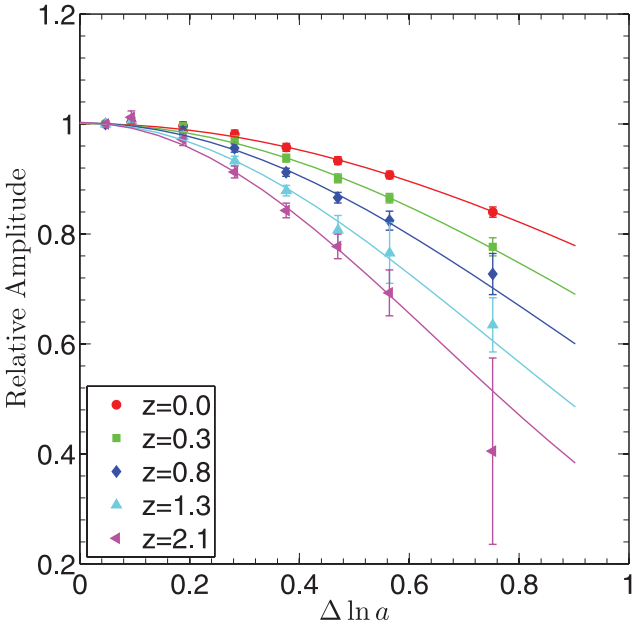

Figure C1. The effect of time resolution on subhalo abundance. Data points are the relative amplitudes of the diluted subhalo mass functions to the best resolved one, as a function of time resolution. We show the result at five different redshifts, together with a Gaussian function fitting at each redshift. Error bars in the figure are propagated from Poisson errors on the subhalo mass functions.

The other aspect of the time resolution problem is whether the unbinding routine would be dependent on the time resolution adopted. We still use the halo S51G86 studied before to test this. As shown in Fig. C2, applying the adaptive unbinding algorithm with different time resolution has no noticeable effect on the resulting self-bound mass. This again demonstrates the robustness of our core-averaged unbinding algorithm in conjugation with our adaptive source management.

APPENDIX D: UNILATERALLY MATCHED SUBHALOES

Unilateral subhaloes in Fig. 8 come in several situations. First there are subhaloes found by one finder but missed by the other, and thus they appear as part of a larger subhalo, either central or satellite. These are reflected as vertical lines connecting up or horizontal lines connecting right to bilateral subhaloes. Most unilateral subhaloes belong to this case, with the majority having their correspondences being the central. For those unilateral HBT subhaloes linked to a satellite, most of them are in fact overlapping with their correspondence, having a separation between their centres being the order of the smoothing length of dark matter particles. In fact if they can keep this overlapping state, these two subhaloes should be treated as one, as merger between them has completed. However, because

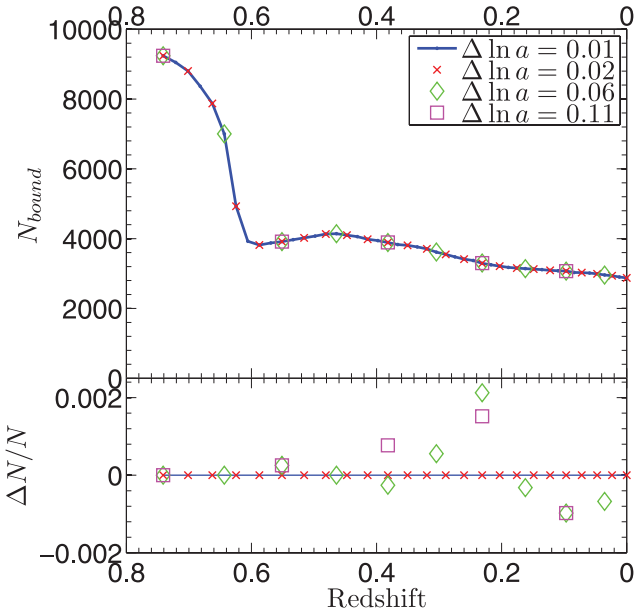

Figure C2. The effect of time resolution on the self-bound mass. We apply the code skipping different number of snapshots at every tracing step, to test the stability of our unbinding algorithm with respect to time resolution. In the top panel, the solid line denotes the bound mass when using all the snapshots available, while different symbols give the results of skipping different number of snapshots, labelled as different steps in the scale factor a. In the lower panel, we show the relative difference with respect to the highest resolution case.

overlapping subhaloes are still rare, and because it is not clear whether these pairs have mixed completely, we do not integrate them together in the current version of HBT. We also note that one SUBFIND subhalo with 2×10^3 particles is missed by HBT, indicating that the tracing algorithm is still not perfect although it is already massively optimized. However, as statistically found in our comparison using the cosmological simulation, this kind of missed subhaloes can be as large as 10^4 particles for SUBFIND, amounting to as high as 10 per cent in number near that mass, while HBT only misses less than 1 per cent for the highest missed mass of 10^3 particles.

A second situation is when one subhalo finder only identifies the most inner part of a subhalo and assigns a majority of surrounding bound particles to the central subhalo, while its correspondence identifies both parts. When doing the match, the surrounding particles from the correspondence outweighs the inner part, linking the correspondence to the central subhalo. This situation can be found as two adjacent perpendicular lines connecting two unilateral subhaloes and the central. It happens for both finders. In the SUBFIND case, it is due to the truncation around the radius of equality as illustrated in Section 3.4 while in the HBT case we attribute this to the effect of local accretion discussed in Appendix B.

The third situation is similar to the second one but happens for binary subhaloes, which has been discussed in the main text as 'sub-in-sub' switch.

APPENDIX E: ACCURACY OF TWO ESTIMATORS FOR THE SUBHALO MASS FUNCTION

The subhalo mass function $f(m) \equiv \mathrm{d}N/\mathrm{d}\ln m$, where m is the subhalo mass (either $M_{\rm sub}$ or $M_{\rm sub}/M_{\rm host}$), can be estimated directly

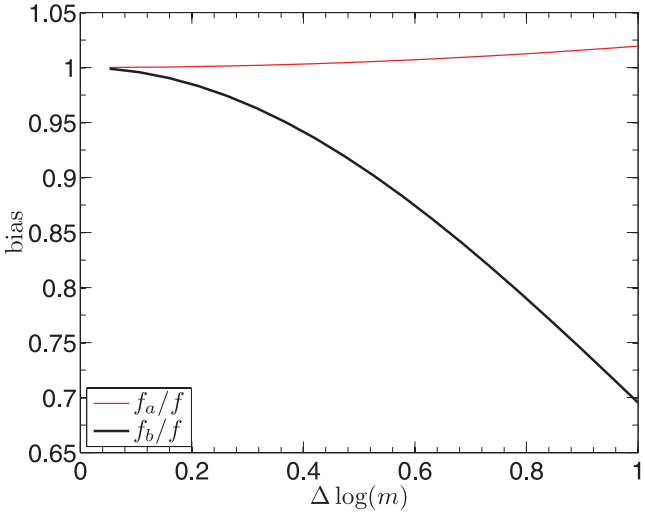

Figure E1. Bias of mass function estimators as a function of mass bin size $\Delta\log(m) = \log(k)$.

from two discrete estimators as

$$f_a(\bar{m}) = \frac{\Delta N}{\Delta \ln m} \tag{E1}$$

or

$$f_b(\bar{m}) = \frac{1}{\Delta m} \sum_{m_1 < m_i < m_2} m_i, \tag{E2}$$

where m_1 and m_2 are the lower and upper limits of the subhalo mass bin, ΔN is the number of subhaloes in the bin and $\bar{m}$ is the average subhalo mass in the bin. In the limit $\Delta m \to 0$, we have $\bar{m} = m$ and $f_a(\bar{m}) = f_b(\bar{m}) = f(m)$. For a finite mass bin, suppose $f(m) = m^{-1}$ ignoring the normalization, then

$$f_a(\bar{m}) = -\frac{\Delta(1/m)}{\Delta \ln m},$$

$$f_b(\bar{m}) = \frac{\Delta \ln m}{\Delta m},$$

$$f(\bar{m}) = -\frac{\Delta(1/m)}{\Delta \ln m}.$$

It is ready to see that the estimator f_a recovers exactly f. For a more general form $f(m) = m^{-\gamma}$, let $k = m_2/m_1$, then the biases of the two estimators are given as

$$f_a(\bar{m})/f(\bar{m}) = \frac{1}{-\gamma} \frac{k^{-\gamma} - 1}{\ln k} \left(\frac{\gamma}{\gamma - 1} \frac{k^{1-\gamma} - 1}{k^{-\gamma} - 1} \right)^{\gamma}, \tag{E3}$$

$$f_b(\bar{m})/f(\bar{m}) = \frac{1}{1 - \gamma} \frac{k^{1-\gamma} - 1}{k - 1} \left(\frac{\gamma}{\gamma - 1} \frac{k^{1-\gamma} - 1}{k^{-\gamma} - 1} \right)^{\gamma}. \tag{E4}$$

Since the biases do not depend on m, the slope of the mass function is recovered for both estimators while the normalizations differ. In Fig. E1, we show the bias of these two estimators for $\gamma = 0.9$. With bin size $\Delta\log(m) = 0.5$, f_b will underestimate the mass function by 10 per cent. Throughout this work, we use f_a to calculate subhalo mass function except in Figs 3 and 5 where f_b is used to give the total subhalo mass contained in each mass bin.

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