An accurate comprehensive approach to substructure: III. Masses and formation times of the host haloes

Eduard Salvador-Solé,[1]★ Alberto Manrique,[1] David Canales[2] and Ignacio Botella[1,3]

[1]*Institut de Ciències del Cosmos, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*
[2]*School of Aeronautics and Astronautics, Purdue University, 701 W. Stadium Ave., West Lafayette, IN 47907, USA*
[3]*Department of Astronomy, Graduate School of Science, Kyoto University, Kitashirakawa, Oiwakecho, Sakyo-ku, Kyoto, 606-8502, Japan*

ABSTRACT

With this paper, we complete a comprehensive study of substructure in dark matter haloes. In Paper I, we derived the radial distribution and mass function (MF) of accreted subhaloes (scaled to the radius and mass of the host halo) and showed that they are essentially universal. This is not the case, however, for those of stripped subhaloes, which depend on halo mass and assembly history. In Paper II, we derived these latter properties in the simplest case of purely accreting haloes. Here, we extend the study to ordinary haloes having suffered major mergers. After showing that all the properties of substructure are encoded in the mean truncated-to-original subhalo mass ratio profile, we demonstrate that the dependence of the subhalo MF on halo mass arises from their mass-dependent concentration, while the shape of the subhalo radial distribution depends on the time of the last major merger of the host halo. In this sense, the latter property is a better probe of halo formation time than the former. Unfortunately, this is not the case for the radial distribution of satellites as this profile is essentially disconnected from subhalo stripping and the properties of accreted subhaloes are independent of the halo formation time.

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

1 INTRODUCTION

In the last decade, the subject of halo assembly history has attracted much attention in connection with the so-called 'missing satellite problem', namely that the abundance of satellite galaxies in the Milky Way (MW) (and Andromeda; Tollerud, Boylan-Kolchin & Bullock 2014) does not seem to conform with the expected one in the favourite ΛCDM cosmology (Klypin et al. 1999; Moore et al. 1999; Bullock & Boylan-Kolchin 2017). Indeed, one possible explanation for that problem is that the MW halo may not have had the typical assembly history of haloes of its mass. Some aspects of the MW suggest, indeed, that it has had a particularly quiescent history (e.g. Wyse 2001; Deason et al. 2013; Ruchti et al. 2015; Lancaster, Belokurov & Evans 2019).

Substructure in dark matter haloes is believed to harbour important information on their assembly history as it is the direct consequence of the way they have grown. Indeed, haloes undergo long periods of frequent *minor* mergers (generically called accretion), separated by sporadic *major* mergers. The difference between these two kinds of mergers is that the largest halo in minor mergers (the accreting object) is much more massive than its partners (the accreted objects), so it remains essentially in equilibrium during that process and the accreted haloes survive within their host as subhaloes. On the contrary, all haloes (usually two) partaking of a major merger are similarly massive so that the event causes them to go out of equilibrium and to form a new virialized halo after the system relaxes

again. In this case, the progenitor haloes are thus destroyed, but their subhaloes are transferred to the newborn halo. Subhaloes thus accumulate within haloes, where they are more or less stripped and shock-heated depending on the characteristics of each transient host. Consequently, their final properties are the result of their past history.

Unfortunately, we do not know yet what are the typical properties of substructure in haloes of different masses and formation times. None of the analytic models of halo substructure so far developed (Taylor & Babul 2001; Sheth 2003; Zentner & Bullock 2003; Lee 2004; Oguri & Lee 2004; Taylor & Babul 2004; Peñarrubia & Benson 2005; van den Bosch, Tormen & Giocoli 2005; Zentner et al. 2005; Kampakoglou & Benson 2007; Giocoli, Tormen & van den Bosch 2008; Benson et al. 2013; Pullen, Benson & Moustakas 2014; Griffen et al. 2016; Jiang & van den Bosch 2016; van den Bosch & Jiang 2016; van den Bosch et al. 2018; Green & van den Bosch 2019; Font et al. 2020) have been able to provide definite answers to these questions, nor have high-resolution simulations, which have only been able to draw the properties of substructure in a handful of haloes of the MW mass (Diemand, Kuhlen & Madau 2007; Springel et al. 2008; Boylan-Kolchin et al. 2010; Wang et al. 2011) or of other masses (e.g. Angulo et al. 2009; Elahi, Widrow & Thacker 2009; Giocoli et al. 2010; Klypin, Trujillo-Gomez & Primack 2011; Gao et al. 2011, 2012; Onions et al. 2012; Cautun et al. 2014; Lovell et al. 2014; Ishiyama et al. 2020; Lovell et al. 2021).

But things are rapidly changing. The incoming new data gathered by means of the *Gaia* satellite (Gaia Collaboration 2018) will allow to accurately determine the MW's substructure and it has now become feasible to observe satellites in neighbouring MW analogues (Danieli et al. 2017; Geha et al. 2017; Smercina et al. 2018; Bennet

et al. 2019, 2020; Crnojević et al. 2019; Carlsten et al. 2020; Mao et al. 2020; Carlsten et al. 2021). In parallel, simulations have also greatly improved. The *Copernicus Complexio N*-body simulations together with a semi-analytic galaxy formation model (Hellwing et al. 2016; Bose et al. 2016, 2020) or the *Apostle* and *Auriga* (Richings et al. 2020), *FIRE-2* (Samuel et al. 2020), and *Artemis* (Font et al. 2020; Engler et al. 2021; Font, McCarthy & Belokurov 2021) hydrodynamic simulations have gathered a considerable number of simulated MW (and Andromeda) analogues. On the other hand, it is now possible to reach very high stellar mass resolutions (Grand et al. 2021) that significantly improve the statistics of substructure at the level of faint and ultra-faint satellites ($10^2 \, \mathrm{M_\odot} < M_\star < 10^6 \, \mathrm{M_\odot}$).

In addition, great progress has also been made on analytic grounds. A very complete model of substructure formation has been built (Jiang et al. 2021) that allows one to study the effects of different initial conditions in the accretion and evolution of satellite galaxies. Likewise, the powerful *ConflUent System of Peak* (*CUSP*) *trajectories* formalism (Manrique & Salvador-Solé 1995, 1996; Manrique et al. 1998; Salvador-Solé et al. 2012a,b; Juan et al. 2014a,b), making the link between the properties of haloes and their seeds (peaks) in the random Gaussian field of density perturbations (see Salvador-Solé & Manrique 2021 for an overview), has been successfully applied to the study of halo substructure (Salvador-Solé, Manrique & Botella 2021a and Salvador-Solé, Manrique & Botella 2021b, hereafter Papers I and II, respectively).

In Paper I, we determined the properties of accreted subhaloes, which act as initial conditions in their evolution through stripping inside their host haloes. To do this, we took profit of the fact that, as shown in Salvador-Solé et al. (), all halo properties arising from their gravitational clustering process do not depend on their particular assembly history so that one has the right to focus on the simplest case of *purely accreting haloes* evolving inside-out (see below for a brief explanation of this important result). However, subhalo stripping and dynamical friction are two (coupled) mechanisms acting on the dynamical evolution of subhaloes that are not directly connected to gravitational clustering. Consequently, the properties of stripped subhaloes do depend on the halo assembly history. In Paper II, we built a detailed model of tidal stripping and shock heating of subhaloes as they orbit inside haloes, which allowed us to derive their final properties. However, this was done in the simplest case of purely accreting haloes only and neglecting dynamical friction.

In the present paper, we complete this study and extend the characterization of substructure to ordinary haloes, i.e. haloes having suffered major mergers, paying special attention to the role of halo mass and formation time and analysing the possible use of those properties as a probe for halo assembly history. Our treatment does not include dynamical friction. However, the predictions for low-mass subhaloes ($M_s \lesssim 10^{-4} \, M_h$, where M_h is the mass of the host halo) should not be affected by that omission. On the other hand, it does not include baryons either. However, by comparing our predictions to the results of simulations including them, it is still possible to unravel to some extent the influence of baryons physics in the properties of substructure.

The layout of the paper is as follows. In Section 2, we remind the main results of Papers I and II for purely accreting haloes of different masses. In Section 3, we extend those results to haloes having suffered major mergers. And in Section 4, we analyse the properties of substructure in haloes of a fixed mass and different formation times. Our results are summarized and discussed in Section 5.

Some comments on the notation used in this Paper are in order. Unless otherwise stated, when we refer to subhaloes without spec-

ifying their kind, we mean stripped (or truncated) subhaloes. The halo formation time used is defined as the time they suffered their last major merger. It thus differs from the most usual definition: the time haloes reach 50 per cent of their final mass. We prefer the former not only because it is less arbitrary (why 50 per cent and not, say, 75 per cent?) but also because it is physically better motivated. Indeed, as mentioned, virialized haloes interrupt their identity in major mergers where they are destroyed and a new virialized object appears (Salvador-Solé & Manrique 2021). Lastly, the notation we use for the cumulative or differential subhalo abundances, dependent in general on subhalo mass, M_s, and radial location inside the host halo, r, is the same as in Papers I and II. The cumulative number of stripped or accreted haloes out to r and down to M_s are denoted as $\mathcal{N}^{\mathrm{stp}}(<r, >M_s)$ and $\mathcal{N}^{\mathrm{acc}}(<r, >M_s)$, respectively. When one of the arguments takes its maximum value, i.e. when the integrals over r or M_s are complete, we drop the corresponding argument. For instance, $\mathcal{N}^{\mathrm{stp}}(>M_s)$ stands for $\mathcal{N}^{\mathrm{stp}}(<R_h, >M_s)$, where R_h is the total halo radius, and $\mathcal{N}^{\mathrm{stp}}(<r)$ stands for $\mathcal{N}^{\mathrm{stp}}(<r, >0)$ [or for $\mathcal{N}^{\mathrm{stp}}(<r, <M_h)$, where M_h is the total halo mass]. Lastly, the differential form with respect to any argument, r or M_s, of any of the previous functions is denoted without the corresponding preceding inequality symbol. For example, $\mathcal{N}^{\mathrm{stp}}(r, M_s)$ stands for the double derivative of the cumulative abundance with respect to r and M_s and $\mathcal{N}^{\mathrm{stp}}(M_s)$ stands for the differential subhalo mass function (MF). This greatly simplifies the notation as it avoids writing the dumb arguments R_h or M_h (or 0) as well as the symbols of single and multiple derivatives in most expressions.

Also like in Papers I and II, M_h is defined as the mass encompassed by the virial radius R_h within which the inner mean density is equal to the virial overdensity (Bryan & Norman 1998; Henry 2000) times the current mean cosmic density. In particular, we assume the MW mass equal to $M_h = 2.2 \times 10^{12} \, \mathrm{M_\odot}$. The cosmology adopted is that given by the best *WMAP7* parameters (Komatsu et al. 2011), with CDM spectrum according to the prescription given by Bardeen et al. (1986) with the Sugiyama (1995) shape parameter. The reader is referred to Papers I and II for the role of diffuse dark matter (dDM) in the properties of substructure, just briefly referred to in this paper.

2 PURELY ACCRETING HALOES

During accretion, haloes evolve inside-out because the later particles (and subhaloes) fall on to them, the larger their initial turnaround radius as well as their final apocentric radius due to the ordered virialization process taking place in this case (Salvador-Solé & Manrique 2021). On the contrary, their structure is completely rearranged in major mergers through violent relaxation. This causes the memory loss of their past history, so that their final properties are indistinguishable from those of purely accreting haloes with the same mass M_h at the same cosmic time t_h. A formal proof for this important result is provided in Salvador-Solé & Manrique (2021), but the origin of it is as follows. There is a one-to-one correspondence between haloes with mass M_h at the cosmic time t_h and their seeds: peaks of density contrast δ in the initial Gaussian random density field filtered with a Gaussian window of scale S. That correspondence does not involve any other halo or peak characteristic. In particular, it does not depend on how clumpy the initial mass distribution is inside the initial patch encompassed by the filter or, equivalently, on how lumpy the collapse of the peak is.

Therefore, all properties of haloes with M_h at t_h arising from gravitational clustering, i.e. from their mass assembly, through accretion and major mergers, are degenerate with respect to their

formation time. This is why to study them one has the right to assume pure accretion. In fact, since gravitation is scale-free, such halo properties would be strictly universal (i.e. independent of halo mass and formation time), except for the scale length introduced by the power spectrum of density fluctuations in the CDM cosmology at the base of the typical mass–concentration M–c relation (see Salvador-Solé et al., in preparation).

However, subhalo stripping is not related to gravitational clustering and its effects on subhaloes are not erased by violent relaxation. Consequently, stripped subhaloes do retain the memory of the halo assembly history. This is why *substructure is expected to depend on halo formation time* and, since haloes with different masses have different typical formation times, *on halo mass as well*. In this sense, the properties of substructure derived in Paper II for purely accreting haloes might substantially differ from the properties of ordinary haloes having suffered major mergers. Nevertheless, to understand the latter, we need first to comprehend the former from which they follow (see Section 3). It is thus worthwhile reminding the results of Paper II.

As explained in Papers I and II, we distinguish between 'the time of accretion of a subhalo' on to the halo and 'the time of its first crossing'. The latter corresponds to the first time the subhalo orbits within the virialized and non-virialized parts of the halo after reaching turnaround. During the first few crossings of the system, subhalo orbits shrink due to their energy exchange with the shells they cross, which causes the non-virialized part of the subhalo to contract adiabatically. But, after these few initial crossings and neglecting the effects of dynamical friction, subhalo orbits stabilize, with the apocentre at the instantaneous virial radius of the newly virialized part of the halo, which thus grows inside-out. The time at which subhalo orbits become stable is what we adopt as the time of their accretion on to the virialized halo.[1] In what follows, we concentrate in monitoring the stripping of subhaloes after their accretion on to the host halo, i.e. once their orbits are fixed (neglecting dynamical friction).

The stripped subhalo abundance per infinitesimal truncated mass and radius at M_s^{tr} and r within a purely accreting halo with M_h at t_h is given by

$$\mathcal{N}^{stp}\left(r, M_s^{tr}\right) = \mathcal{N}^{tr}\left(r, M_s^{tr}\right) + \left\langle \int_{M_s}^{M(r)} dM\, \mathcal{N}^{acc}(r, M) \right.$$
$$\left. \times \int_{R^{tr}(v,r,M)}^{R(r,M)} dr'\, \mathcal{N}_{[M,t(r)]}^{stp}\left(r', M_s^{tr}\right) \right\rangle, \quad (1)$$

where angular brackets indicate average over the tangential velocity v of subhaloes at their apocentre at r where they spend most of the time. The subindex [M,t] in the properties of subhaloes is to indicate that they refer to host haloes with M at t. For simplicity in the notation, we have skipped the subindex [M_h,t_h] for the host halo itself, but we will re-introduce it in Sections 3 and 4 when dealing with haloes of different masses and times.

The first term on the right of equation (1), equal to

$$\mathcal{N}^{tr}\left(r, M_s^{tr}\right) = \mu\left(r, M_s^{tr}\right) \mathcal{N}^{acc}\left(r, M_s^{tr}\right), \quad (2)$$

gives the contribution directly arising from the stripping of accreted subhaloes of suited mass M_s, which have their apocentre at r (see Paper I) and whose abundance is

$$\mathcal{N}^{acc}(r, M_s) = 4\pi r^2 \frac{\rho(r)}{M_h} \mathcal{N}^{acc}(M_s), \quad (3)$$

where $\mathcal{N}^{acc}(M_s)$ is their differential MF and $\rho(r)$ is the halo density profile. In equation (2), $\mu(r, M_s^{tr})$ is the mean (averaged over v) truncated-to-original mass ratio of subhaloes with M_s^{tr} at r, calculated in Paper II by monitoring the mass loss through repetitive stripping and shock heating of subhaloes accreted at $t(r)$ when the host halo had radius r and mass $M(r)$. And the second term on the right of equation (1) gives the contribution arising from subsubhaloes that were lying in accreted subhaloes with mass M and tangential velocities v at r and have been released into the intra-halo medium when their hosts, with initial radius $R(r, M)$, have been truncated at $R^{tr}(v, r, M)$. See Paper II for the expression of the truncation radius in objects with the NFW (Navarro, Frenk & White 1997) density profile.

Equation (1) can be rewritten in the simple form

$$\mathcal{N}^{stp}(r, M_s^{tr}) = \left[1 + f_{rel}\left(r, M_s^{tr}\right)\right] \mu\left(r, M_s^{tr}\right) \mathcal{N}^{acc}\left(r, M_s^{tr}\right), \quad (4)$$

where $f_{rel}(r, M_s^{tr})$ is the virtual fraction of accreted subhaloes with mass M_s^{tr} converted into stripped ones of that mass at r arising from subsubhaloes. This fraction is the solution of the Fredholm integral equation of the second kind (with the boundary condition $f_{rel}(0, M_s^{tr}) = 0$) of the differential equation

$$-\frac{d}{dr}\left[f_{rel}\left(r, M_s^{tr}\right) \mu\left(r, M_s^{tr}\right) \frac{R_h^3}{r^3}\right] = \left\langle \frac{\partial \mathcal{R}}{\partial r} \frac{\mathcal{N}^{stp}\left(\mathcal{R}, M_s^{tr}\right)}{\mathcal{N}^{acc}\left(M_s^{tr}\right)} \right\rangle, \quad (5)$$

with $\mathcal{R}(v, r) \equiv R^{tr}(v, r, M_s) R_h / R_s(r, M_s)$. The function f_{rel} is always less than a few percent, so it can be safely neglected in front of unity, though, for the sake of completeness, it is kept in all the expressions below. On the contrary, the mean truncated-to-original subhalo mass ratio, $\mu(r, M_s^{tr})$, depicted in Fig. 1 plays a crucial role in the properties of substructure. It is thus worth explaining its main features.

As the strength of stripping and shock heating depends on the concentration c of both subhaloes and the host halo (see Paper II) and c depends on the mass of the object through the well-known mass–concentration (M–c) relation, μ depends on the masses of subhaloes and the host halo. Although those dependencies are not gathered in Fig. 1, which focuses on the μ profile for subhaloes with 10^8 M$_\odot$ in haloes with 2.2×10^{12} M$_\odot$, they are very important to understand the properties of substructure. Indeed, the μ varies in a non-trivial way according to the mass of the host halo: at R_h, it is always close to unity, but its inwards decrease is less steep in more massive haloes because their concentration is lower, which causes the pericentric radius reached by subhaloes with identical v at apocentre to be larger so that stripping and shock heating are less marked (see Paper II). As a result, the mass integral of μ varies with halo mass as $\propto M_h^{0.08}$. Regarding the dependence of μ on subhalo mass, it turns out that $\mu(r, M_s^{tr})$ is separable. The reason for this is that subhaloes are truncated by tidal stripping at the radius (dependent on their own concentration) where the inner mean density equals that of the halo at the pericentre, which is the same for subhaloes of all masses with identical v at r (see Paper II). The factor dependent on subhalo mass is essentially proportional to $(M_s^{tr})^{-0.03}$. Both mass dependencies are weak, however, particularly that on subhalo mass due to the fact that subhaloes accreted at any time $t(r)$ have similar concentrations. In this sense, the μ profile for any fixed halo mass can be seen, in a first approximation, to depend only on r.

---

[1] Strictly speaking, the continuous arrival of new subhaloes that cross the virialized halo causes it a slight adiabatic contraction. But the characteristic time-scale of this effect is very long and it can be safely ignored. This is in fact the reason why we can see the inner halo as virialized despite the continuous non-stabilized recent arrivals crossing it.
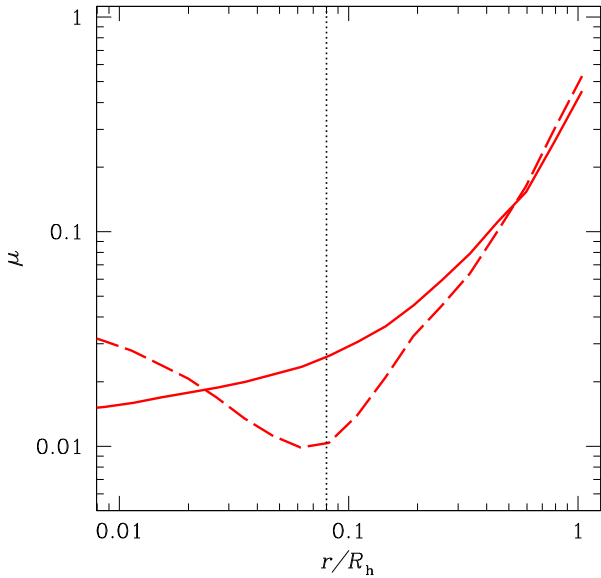
Figure 1. Mean truncated-to-original subhalo mass ratio profile predicted by CUSP for subhaloes with $M_s = 10^8$ M$_\odot$ in purely accreting MW–mass haloes using the unbiased CUSP M–c relation (solid red line) and the Gao et al. ([2008](#)) empirical M–c relation affected by the limited mass resolution of simulations (long-dashed red line). The vertical dotted black line marks the radius where the inwards decreasing behaviour of μ in the latter case is inverted for the reason explained in the text.
(A colour version of this figure is available in the online journal.)
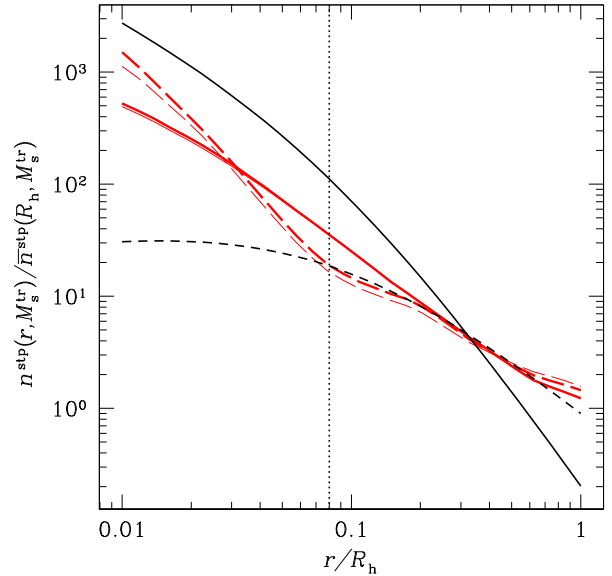

Figure 2. Scaled number density profiles of subhaloes with different masses (red lines) predicted for purely accreting haloes with current MW–mass using the CUSP M–c relation (solid red lines) and the Gao et al. ([2008](#)) M–c relation affected by the limited mass resolution of simulations (long-dashed red lines). For comparison, we plot the fit by Han et al. ([2016](#)) to the profile found in the halo A of the Level 1 Aquarius simulation (black dashed line) affected by a similar resolution but having not evolved by accretion before $z \sim 6$ corresponding to the radius marked with a vertical dotted black line like in Fig. [1](#). The solid black line is the scaled halo density profile. To better appreciate the effect of changing the subhalo masses, we plot the predictions for subhaloes with $10^{-2}M_h$ and $10^{-4}M_h$ in thick and thin red lines, respectively.
(A colour version of this figure is available in the online journal.).

A more subtle issue is that, for the above-mentioned reasons, the shape of the μ profile will depend on the M–c relation. Of course, the real μ profile predicted by CUSP implicitly follows from the M–c relation that can be derived within that framework (Salvador-Solé et al. in preparation). However, if we want to reproduce the results of simulations, we must use the M–c relation found in numerical studies by, e.g. Gao et al. ([2008](#)) with a limited mass resolution similar to that affecting those empirical results. The effect of the limited mass resolution of simulations, which affects all (sub)haloes at early times when they are little massive, is apparent in Fig. [1](#). While the μ profile derived from the CUSP M–c relation with no limited mass resolution is ever decreasing inwards, that found with the Gao et al. ([2008](#)) M–c relation stops decreasing at $r \sim 0.08R_h$ and then begins to increase again. The reason for this strange result is the following. In a purely accreting halo evolving inside-out as considered here, its concentration at earlier times decreases with decreasing r as r/r_s, because of the fixed value of r_s. Since the concentration of accreted subhaloes also decreases with increasing z (this is so in all empirical as well as theoretical M–c relations), stripping keeps on being effective at small radii populated by subhaloes accreted at higher redshifts. However, due to the limited mass resolution, the concentration of (sub)haloes in the Gao et al. ([2008](#)) M–c relation is bounded to a minimum value independent of mass reached at $z \sim 3$ ($r \sim 0.3R_h$). As a consequence, at radii smaller than $0.3R_h$, the concentration of the inside-out evolving host halo continues to decrease, while that of subhaloes does not, and subhalo stripping becomes (artificially) ineffective.

The previous discussion also illustrates that, in normal conditions, the μ profile decreases inwards despite the fact that the smaller the radius, the lower the concentration of the host halo seen by subhaloes. Indeed, the main cause shaping the μ profile is the time subhaloes

at different radii have been undergoing stripping rather than the different typical tidal forces they see.

Since both $\mu(r, M_s^{tr})$ and $\mathcal{N}^{acc}(r, M_s^{tr})$ are separable, so is also the radial abundance of stripped subhaloes $\mathcal{N}^{stp}(r, M_s^{tr})$ [equation (4)]. Thus, the number density profile per infinitesimal mass of stripped subhaloes, $n^{stp}(r, M_s^{tr}) \equiv \mathcal{N}^{stp}(r, M_s^{tr})/(4\pi r^2)$, scaled to their total number in the halo,

$$\frac{n^{stp}(r, M_s^{tr})}{\bar{n}^{stp}(R_h, M_s^{tr})} = \frac{\mathcal{N}^{stp}(r, M_s^{tr})/(4\pi r^2)}{3\mathcal{N}^{stp}(M_s^{tr})/(4\pi R_h^3)}, \tag{6}$$

is independent of subhalo mass. From now on, a bar on a quantity dependent on r means the radial average of that quantity inside r.

In Fig. [2](#), we show the scaled number density profiles of subhaloes of two masses ($10^{-2}M_h$ and $10^{-4}M_h$) found for the two above-mentioned M–c relations. For comparison, we plot the profile found by Han et al. ([2016](#)) in the MW–mass halo A of the Level 1 *Aquarius* simulation (Springel et al. [2008](#)), also found to be roughly independent of subhalo mass (except for the effects of dynamical friction; Han et al. [2018](#)). This halo is particularly well suited to the comparison with our predictions for purely accreting haloes because it suffered the last major merger at $z \sim 6$ ($r \sim 0.08R_h$) and has been evolving by accretion (and growing inside-out) since then. Note that, given its formation time, the concentration r/r_s cannot be traced down to radii smaller than $r = 0.08R_h$. This is the reason that its μ profile is not seen to increase at smaller radii. In fact, as we will see in Section 3, it should be flat there, though the μ profile of the simulated
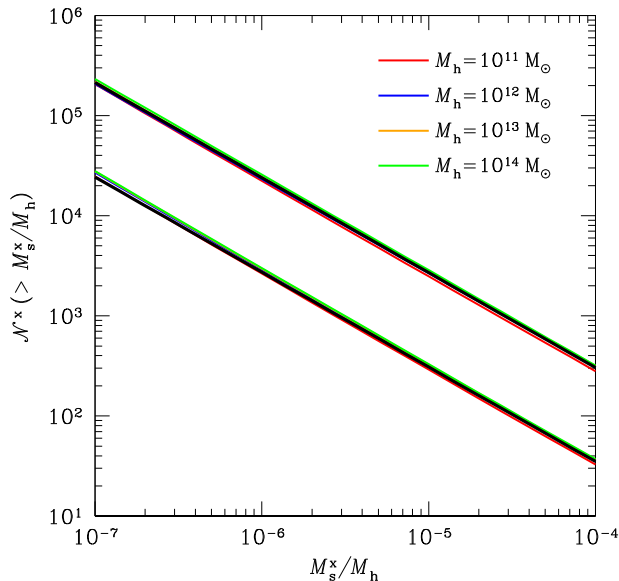
Figure 3. Cumulative MF of *accreted* subhaloes (upper curves) and of *stripped* subhaloes (lower curves) (superindex 'x' on $\mathcal{N}$ stands for 'acc' or 'str', respectively) predicted by CUSP for purely accreting haloes of the different quoted masses M_h. At the scale of the plot, the predictions for the Gao et al. or CUSP M–c relations coincide. The MFs of accreted subhaloes for haloes with different masses overlap as their counterparts in simulated ordinary haloes (solid black lines) and the same is true for the MFs of stripped subhaloes provided they are multiplied by $[M_h/(10^{12}\ M_\odot)]^{-0.08}$. These results agree with what is found in simulated ordinary haloes (solid black lines) as derived by Han et al. 2018, here properly normalized so as to include first-level subhaloes only (see Paper II).
(A colour version of this figure is available in the online journal.)

halo is not well determined at those radii because it is dominated by orphan objects.[2]

Like Figs 1 and 2 focuses on MW–mass haloes, so it does not inform on the dependence on halo mass of the scaled subhalo number density profile in purely accreting haloes. However, our calculations show that the profiles for haloes of different masses are quite similar. They do not overlap, however, because, as mentioned, the corresponding μ profiles are not simply shifted with respect to each other by a constant factor (see Fig. 6 dealing with ordinary haloes).

Lastly, integrating over r the subhalo abundance given in equation (4), we obtain the subhalo differential MF, which takes the form

$$\mathcal{N}^{stp}\left(M_s^{tr}\right) = \overline{(1+f_{rel})\mu}(R_h)\,\mathcal{N}^{acc}\left(M_s^{tr}\right). \tag{7}$$

And integrating it over subhalo mass from M_s^{tr}, we arrive at the cumulative MF, $\mathcal{N}^{stp}(>M_s^{tr})$.

In Fig. 3, we plot the cumulative MFs of accreted and stripped subhaloes predicted for purely accreting haloes. As can be seen, the MF of accreted subhaloes is universal, i.e. independent of halo mass, in agreement with the results of simulations (Han et al. 2018 and references therein). This is equivalent to say that the differential scaled subhalo abundance $\mathcal{N}^{acc}(M_s/M_h)$ per infinitesimal M_s/M_h is also universal or that the subhalo abundance $\mathcal{N}^{acc}(M_s/M_h)$ per infinitesimal M_s varies with halo mass as M_h^{-1} (not to mix up with the differential subhalo abundance $\mathcal{N}^{acc}(M_s)$ per infinitesimal M_s, which

is roughly proportional to M_h; e.g. Fakhouri, Ma & Boylan-Kolchin 2010). The subhalo abundance $\mathcal{N}^{acc}(M_s/M_h)$ per infinitesimal M_s, which will have important consequences in Section 4. In Fig. 3, we also see that the scaled MF of stripped subhaloes is also universal *provided it is multiplied by* $M_h^{-0.08}$. Such a dependence on halo mass of the MF of stripped subhaloes following from the above-mentioned dependence of the integral over r of their μ profile as a consequence of the mass dependence of halo concentration fully agrees with that observed in simulations (the factor rendering the MF of simulated haloes universal is M_h^η with $\eta = -0.1$; Rodríguez-Puebla et al. 2016; Han et al. 2018; see also Zentner et al. 2005; Giocoli et al. 2008; Gao et al. 2011). Strictly speaking, the subhalo MF found in simulations is influenced at the high-mass end by the effects of dynamical friction, ignored in our model. But this effect is insignificant for subhaloes less massive than $10^{-4}M_h$ as represented here (see fig. 10 in Paper II). There is one caveat, however, in this agreement: while our predictions are for purely accreting haloes, the results of simulations refer to *ordinary haloes having suffered major mergers*. In other words, it is not clear whether the agreement is accidental or it will persist when dealing with ordinary haloes.

The empirical result that less massive haloes are poorer (in the sense that their subhalo MF is lower) than more massive ones is commonly interpreted as due to the different formation times of haloes of different masses. Indeed, the less massive a halo, the earlier it typically forms, so: (1) stripping has more time to proceed and (2) it was more efficient because haloes are denser at high redshifts. However, this explanation is at odds with the recent finding by Bose et al. (2020) that the earlier MW–mass haloes form, the richer they are. On the other hand, it is not supported either by the fact that purely accreting haloes show the same dependence despite they all have the same arbitrarily small formation time. Our results rather point to the fact that such a dependence is due to the mass dependence of halo concentration, though we must first confirm that the mass dependence of the MF in purely accreting haloes is preserved in ordinary ones. In fact, our predictions also show that the longer subhaloes have been stripped, the lower their final μ profile. In other words, the different typical concentrations and typical formation times of haloes of different masses go in the opposite direction, so we must clarify which is the dominant effect and why in the richness of ordinary haloes.

3 ORDINARY HALOES OF DIFFERENT MASSES

The density, accreted dDM mass fraction, and mean abundance of accreted subhaloes profiles, $\rho(r)$, $f_{dDM}^{acc}(r)$, and $\mathcal{N}^{acc}(r, M_s)$, respectively, are not related to tidal stripping,[3] so they are the same in both purely accreting and ordinary haloes having suffer major mergers (Salvador-Solé & Manrique 2021). Thus, the only functions in the expression of the radial distribution of stripped subhaloes [equation (4)] that depend on stripping are μ and f_{rel} defined in equations (2) and (5), respectively. Therefore, to obtain the radial distribution of stripped subhaloes in ordinary haloes, we must first determine these two functions in such haloes.

---

[2]Accreted subhaloes are dubbed 'orphan' when their stripped subhaloes have masses below the mass resolution of the simulation.

[3]Stripping redistributes the dDM lost by subhaloes in a leading arm and a trailing tail over their orbits. But, for the same reason that the contribution from subhaloes to the halo density profile can be calculated assuming they lie at their apocentre where they spend most of the time, the contribution from the stripped dDM can also be calculated assuming that it is located at the apocentre of their orbit. Consequently, stripping does not essentially alter the halo density profile.

Next, we show how to obtain these properties in ordinary haloes from their counterparts in purely accreting ones derived in Section 2, hereafter distinguished with superindex PA. We will also use from now the scaled arguments $x = r/R_h$ and $m = M_s/M_h$ so that all functions of those arguments should essentially coincide for purely accreting haloes of different masses. Note that the properties of ordinary haloes of any given mass derived next are their average over all formation times of such haloes. Of course, those theoretical properties are hard to compare to the results of simulations that at present provide the properties of substructure in a small number of haloes of all masses. Nevertheless, these mean properties allow us to elucidate the origin of their dependence on halo mass without being disturbed by the statistical deviations of individual objects.

Be $F_{[M_h, t_h]}(x, m)$ the fraction of accreted subhaloes per infinitesimal mass and radius that satisfy some condition in haloes of M_h at t_h averaged over their formation time (i.e. the time of their last major merger) and $F^{PA}_{[M_h, t_h]}(x, m)$ its counterpart in purely accreting haloes. Taking into account the inside-out growth of haloes after their last major merger, we have the following relation between the two quantities

$$F_{[M_h, t_h]}(x, m) = \int_0^{t(x)} dt \, f_{[M_h, t_h]}(t) \, F^{PA}_{[M_h, t_h]}(x, m)$$
$$+ \int_{t(x)}^{t_h} dt \, f_{[M_h, t_h]}(t) \, \bar{F}_{[M(t), t]}(1, m), \qquad (8)$$

where $f_{[M_h, t_h]}(t)$ is the formation time probability distribution function (PDF) of haloes with M_h at t_h, calculated within the CUSP formalism in Manrique et al. (1998) (see also Raig, González-Casado & Salvador-Solé 2001 for a practical approximate expression in the extended Press-Schechter (EPS) formalism; Press & Schechter 1974; Bond et al. 1991; Bower 1991; Lacey & Cole 1994). $\bar{F}_{[M(t), t]}(x, m)$] stands for the mean fraction $F_{[M(t), t]}(x, m)$] inside x of accreted subhaloes, but, given the form of $\mathcal{N}^{acc}(x, m)$ (equation [3]), it coincides with the simple radial average of $F_{[M(t), t]}(x, m)$ inside x. This is why we denote it with a bar. Note that the contribution on $F(x)$ from haloes formed after $t(x)$ (the second term on the right) is averaged over their own formation times. It thus takes into account the different weight of haloes formed at $t(x)$ with previous different formations times (and so on so forth). On the contrary, the contribution on $F(x)$ from haloes formed before $t(x)$ (the first term) does not depend on their individual formation times (F^{PA} does not depend on any formation time) because, at the radius x, such haloes are accreting material ex novo.

To write equation (8), we have taken into account that, when a halo suffers a major merger, its content is scrambled,[4] so the fraction $F_{[M(t), t]}(x, m)$ at any radius x equals its mean value within the total radius at that moment, $\bar{F}_{[M(t), t]}(1, m)$. In addition, we have taken into account that, after the last major merger, haloes evolve by pure accretion, so $M(t)/M_h$ is the mass track of purely accreting haloes with boundary condition $M(t_h)/M_h = 1$. Note that, even if $F^{PA}_{[M_h, t_h]}(x, m)$ (in scaled arguments) is essentially universal, $F_{[M_h, t_h]}(x, m)$ will depend on M_h and t_h through the explicit dependence on these quantities of the halo formation time PDF (Fig. 4).

Multiplying equation (8) by $\mathcal{N}^{acc}(x, m)$ and integrating over x out to 1, we are led, by partial integration and taking into account
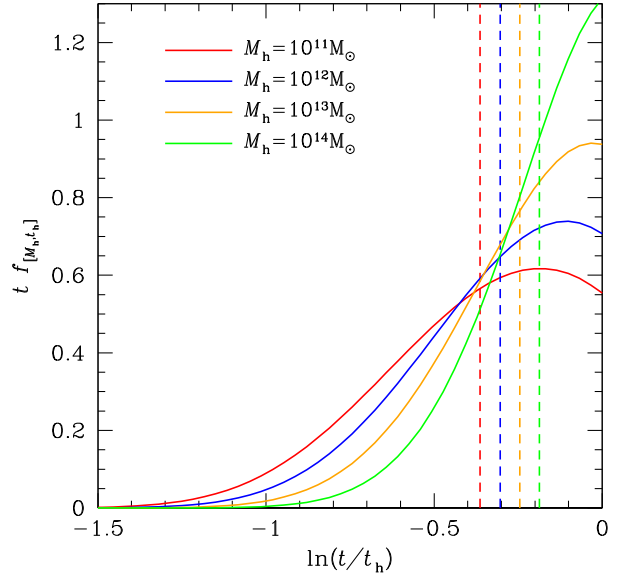
Figure 4. Formation time PDFs of haloes with the quoted masses (coloured lines) at the present time $t_h = t_0$. The coloured dashed vertical lines mark the median formation time t_{med} of haloes of each mass.
(A colour version of this figure is available in the online journal.)

equation (3), to

$$\bar{F}_{[M(t_h), t_h]}(1, m) = \overline{F^{PA}_{[M_h, t_h]} f^c_{[M_h, t_h]}}(1, m)$$
$$+ \int_0^{t_h} dt \, f_{[M_h, t_h]}(t) \, \bar{F}_{[M(t), t]}(1, m) \frac{M(t)}{M_h}, \qquad (9)$$

where $f^c_{[M_h, t_h]}(x)$ stands for the cumulative formation time PDF of haloes with M_h at t_h up to $t(x)$. Equation (9) is a Volterra equation of second kind for $\bar{F}_{[M(t), t]}(1, m)$ as a function of t. Note that, according to equation (9), the mass average of F out to 1, $\bar{F}_{[M(t), t]}(1, m)$, is different from $\bar{F}^{PA}[M(t), t](1, m)$. The reason for this will be seen below. Bringing the solution of this Volterra equation in the integral on the right of equation (8), we arrive at the desired function $F_{[M(t_h), t_h]}(x, m)$ for any value of x (and m). Note also that, as $n^c_{[M_h, t_h]}(x)$ does not depend on m, equations (9) and (8) imply that $F_{[M(t_h), t_h]}(x, m)$ would be independent of m provided its counterpart F^{PA} were. (This is approximately the case for the μ and f_{rel} functions; see next.)

This procedure can be applied to the virtual fractions of accreted subhaloes with M_s^{tr} at r converted into stripped subhaloes by direct stripping, and through the release of subsubhaloes, we arrive at the $\mu(x, m)$ and $f_{rel}(x, m)$ profiles in ordinary haloes of any mass averaged over their formation times. In Fig. 5, we depict the $\mu(x, m)$ profile of ordinary haloes averaged over their formation times that is predicted for the CUSP M–c relation. To better realize the effect of that average, we also plot the schematic (approximate) μ profiles of two individual haloes of 10^{12} M$_\odot$, one formed at $t_f \sim 2$ Gyr (corresponding to $r = 0.2R_h$ in the final halo grown inside-out since that moment) and the other one formed at $t_f \sim 10$ Gyr (corresponding to $r = 0.8R_h$). For the reason mentioned above when explaining the meaning of $\bar{F}_{[M(t), t]}(1, m)$ in the second integral on the right of equation (8), the scrambling of the system at the last major merger yields a flat μ profile equal to its radial average within the formation radius. This is not only the case just after the merger but also long time after. Indeed, after the major merger, all subhaloes inside the scrambled region orbit and suffer stripping during the same time interval $t_h - t_f$ regardless of their past history, so μ in that

---

[4]The scrambling must be complete; otherwise, major mergers would not cause haloes to fully lose the memory of their past history as they do (Salvador-Solé & Manrique 2021).
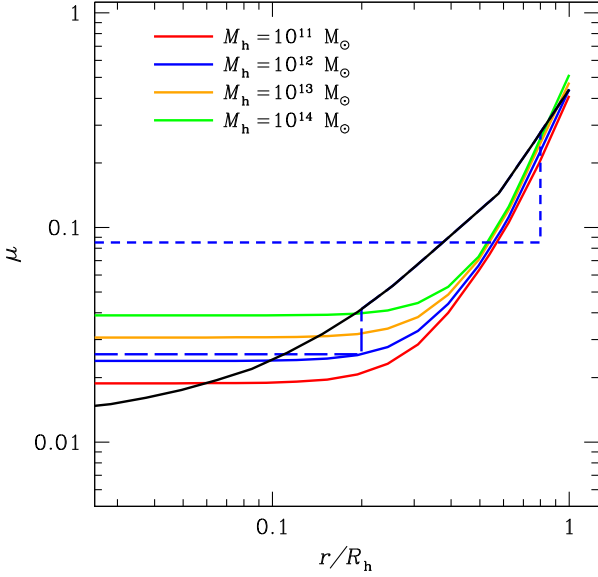
Figure 5. Mean truncated-to-original subhalo mass ratio profiles predicted using the CUSP M–c relation for ordinary haloes of several masses M_h at the current time t_h averaged over their formation times (solid coloured lines). The results for different subhalo masses M_s^{tr} (from 10^6 M$_\odot$ to 10^9 M$_\odot$) overlap when they are multiplied by $[M_s^{tr}/10^8 \text{ M}_\odot]^{-0.03}$. For comparison, we plot the prediction for purely accreting MW–mass haloes (solid black line). To illustrate the effects of averaging over halo formation times, we also plot in long-dashed and short-dashed lines the profiles for two *individual* haloes of 10^{12} M$_\odot$ formed at a high- and low-redshift, respectively.
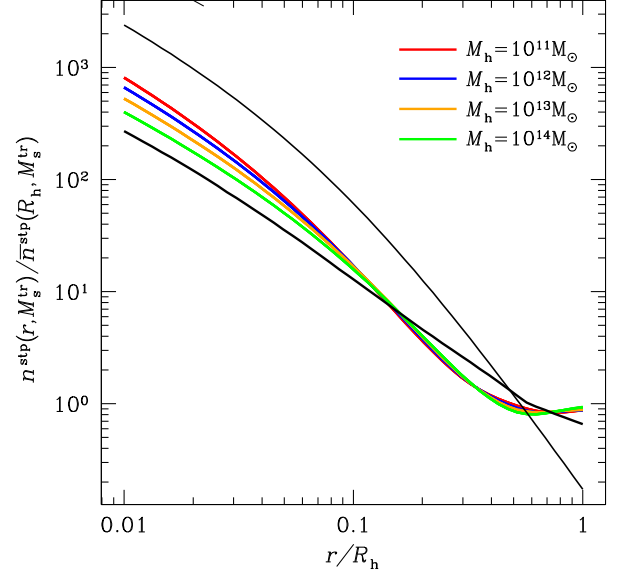(A colour version of this figure is available in the online journal.)

Figure 6. Scaled number density profiles of subhaloes of different masses M_s, which overlap with no added factor, in ordinary haloes of several masses M_h (coloured lines), compared to the profile obtained in purely accreting haloes of 10^{12} M$_\odot$ (thick black line) and the scaled mass density profile of such haloes (thin black line).
(A colour version of this figure is available in the online journal.)

region deepens, but it remains essentially flat until the next merger or the final time. Strictly speaking, subhaloes at different radii within that region suffer different tidal forces over their orbits (the smaller the radius, the lower the concentration of the halo they see). But, as mentioned in connection with Fig. 1, this has a much less marked effect on μ than the different time subhaloes suffer stripping. In fact, if the ending μ profile within that region did show a substantial dependence on radius, the accurate μ profile of haloes averaged over their formation times commented below would show it, which is not the case (see Fig. 5). Likewise, the flat μ profile of those individual haloes inside the formation radius has been taken equal to the radial average inside that radius of the μ^{PA} at the ending time, while it is actually somewhat lower due to the stripping suffered by subhaloes during the time elapsed since the scrambling. Besides those simplifications, the important point to retain from those schematic examples is that, since after the merger haloes evolve inside-out by accretion, their (essentially flat) μ profile in the inner region jumps at its edge to the μ^{PA} profile of purely accreting haloes with the same mass M_h at t_h.

That behaviour of the μ profile in individual haloes of a given mass translates into their formation-time average. This is the reason why the formation-time-averaged profile is also flat at small radii and begins to increase at some radius, dependent on the typical formation time of haloes of that mass, towards the profile of purely accreting haloes. The difference is that, while the profiles of individual haloes reach the purely accreting solution right at the formation radius, the formation-time-averaged profile reaches it only at R_h so that it stays systematically below that solution at large radii. The reason for this difference is clear. Since the μ profiles are outwards increasing, their mass average inside any formation radius is always lower than

the original value at that radius, so the formation-time-averaged profile is also lower than the purely accreting solution. And, as large radii contribute the most to the mass average of μ inside R_h, that mass average in ordinary haloes averaged over their formation times will always be somewhat smaller than that of purely accreting haloes. In any event, the difference should be similar for haloes of any mass, which implies that the mass average of μ in ordinary haloes will essentially coincide with that in purely accreting ones arising, as mentioned in Section 2, from the mass dependence of halo concentration. All these conclusions referring to the μ profiles of ordinary haloes will translate into their subhalo radial distributions and MFs, which will explain their behaviour.

Once the functions μ and f_{rel} have been determined, we can proceed to derive the subhalo abundance per infinitesimal mass and radius around r and M_s^{tr}, $\mathcal{N}^{stp}(r, M_s^{tr})$, in ordinary haloes by application of equation (4). Specifically, since the abundance of accreted subhaloes does not depend on the halo formation time and f_{rel} is negligible, the subhalo radial abundance in ordinary haloes of a given mass averaged over their formation times is simply equal to the formation-time-averaged μ profile times the abundance of accreted subhaloes.

The first consequence of the form of the radial abundance of subhaloes in ordinary haloes refers to the scaled subhalo number density profile (equation [6]). As can be seen in Fig. 6, the profiles of subhaloes of different masses overlap because their μ profiles differ, as mentioned, by the same constant factor as in purely accreting haloes, which cancels with their scaling. Also, like in purely accreting haloes, the profiles for haloes of different masses do not overlap because their μ profiles do not differ by just a constant factor. But there is one interesting difference in comparison with the case of purely accreting haloes: the scaled number density profiles are substantially steeper now. In fact, at small enough radii they are parallel to the halo density profile. (To avoid crowding, in Fig. 6, we plot only the mass density profile and the scaled subhalo number density profile for haloes of 10^{12} M$_\odot$, so that this comparison is
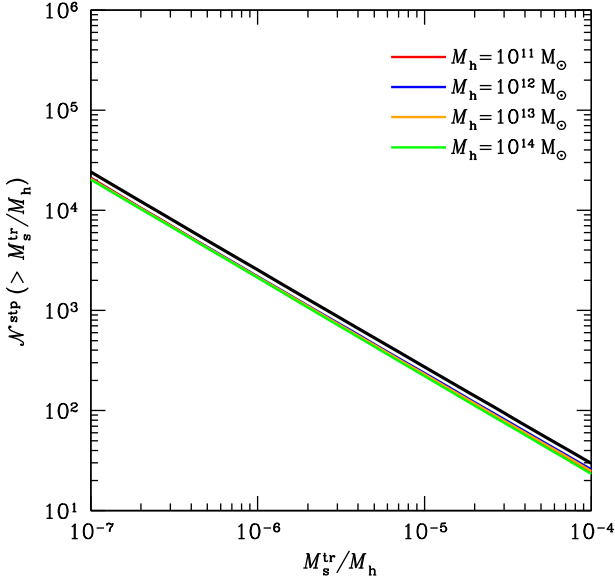
Figure 7. Cumulative subhalo MFs predicted for ordinary haloes of several masses multiplied by $[M_h/(10^{12} \, M_\odot)]^{-0.07}$. The universal MF of purely accreting haloes (thick black line) is just slightly higher.
(A colour version of this figure is available in the online journal.)

possible only for such haloes.) The reason for such a behaviour of the scaled number density profile of ordinary haloes is that in the scrambled regions, subhaloes of all masses have been mixed up with the dDM, so their number density profiles become proportional to the mass density profile of the halo there.

By integrating over r the previous radial abundance of subhaloes, we are led to the differential subhalo MF in ordinary haloes of different masses (equation [7]) and by integration over M_s^{tr} the corresponding cumulative MFs shown in Fig. 7. For the reason mentioned when describing the mass integral of μ, the cumulative MF of ordinary haloes averaged over their formation times shows the same dependence on halo mass as purely accreting haloes. (Strictly speaking, they overlap now when they are multiplied by $[M_h/(10^{12} \, M_\odot)]^{-0.07}$. The slight difference in the power index with respect to that of purely accreting haloes (-0.08) arises from the distinct formation time PDFs of haloes with different masses. In addition, the MF of ordinary haloes averaged over their formation times is somewhat smaller than that of purely accreting haloes, as expected from the discussion on the values of μ near R_h. On the contrary, the MF of individual ordinary haloes fully overlaps with the MF of purely accreting haloes of the same mass, which explains the result in Fig. 3. Only the MF of extreme late-forming haloes is substantially lower because of the rapid decrease of their μ profile when going away from R_h. We will come back to this behaviour of the MF of the latest-forming haloes in the next section.

4 ORDINARY HALOES OF A FIXED MASS AND DIFFERENT FORMATION TIMES

The fraction of accreted subhaloes satisfying any desired property in haloes with a fixed mass M_h at t_h formed in an interval Δ of time around any desired value t_f can be derived using equation (8) with the formation time PDF restricted within that interval, that is

$$f_{[M_h,t_h]}^{t_f,\Delta}(t) = A \, \Pi(t - t_f, \Delta) \, f_{[M_h,t_h]}(t), \tag{10}$$

where $\Pi(t - t_f, \Delta)$ is the top-hat function of width Δ around t_f and A is the normalization factor. This way, we are led to

$$F_{[M_h,t_h]}^{t_f,\Delta}(x, m) = \int_0^{t(x)} dt \, f_{[M_h,t_h]}^{t_f,\Delta}(t) \, F_{[M_h,t_h]}^{PA}(x, m)$$
$$+ \int_{t(x)}^{t_h} dt \, f_{[M_h,t_h]}^{t_f,\Delta}(t) \, \bar{F}_{[M(t),t]}[1, m], \tag{11}$$

Thus, equation (11) with the function $\bar{F}_{[M(t),t]}[1, m]$ solution of the integral equation (9) leads to the desired fraction $F_{[M_h,t_h]}^{t_f \pm \Delta/2}(x, m)$.

This procedure can be applied to infer the μ and f_{rel} profiles in ordinary haloes with formation times averaged inside any desired interval. Following Bose et al. (2020), who studied MW–mass haloes assuming the *WMAP7* cosmology like we do, we consider two extreme intervals: one with an upper bound at $t/t_h = 0.28$ ($z = 1.74$), corresponding to $r/R_h = 0.36$, that embraces the 20 per cent earliest forming objects and another interval with a lower bound at $t/t_h = 0.76$ ($z = 0.29$), corresponding to $r/R_h = 0.82$, that embraces the 20 per cent latest-forming objects. (The previous figures correspond to haloes with $M_h = 10^{12} \, M_\odot$; for haloes of different masses, we have chosen suited values to delimit the same kind of early- and late-forming objects.) We remark that these two intervals coincide with those used by Bose et al. (2020) despite the different halo formation time definition they adopt: the time the halo reaches 50 per cent of its final mass. Indeed, haloes with M_h at t_h that follow the pure accretion track reaches half the final mass at $z = 3.5$. Thus, all haloes undergoing the last major merger after that redshift automatically reach that pure accretion track at the same moment and, hence, they are also seen to form there according to the alternate formation time definition. Haloes undergoing the last major merger before $z = 3.5$ in the alternate formation time definition will be seen, in our definition, to form some time after when they will reach $z = 3.5$. But, since this redshift is higher than the upper redshift of our interval of the earliest forming haloes, all these haloes will lie in that interval according to both formation time definitions, even though their individual formation times will differ in both cases.

The μ profiles for ordinary haloes formed in those two extreme intervals are shown in Fig. 8. Their shape in each case is readily understood from our previous explanations of the behaviour of that profile in Section 3. It also explains the behaviour of the corresponding scaled subhalo number density profiles, $n(r, M_s)/\bar{n}(M_s)$ (equation [6]), shown in Fig. 9, that follow from their radial abundances given by equation (4). As can be seen, the mean scaled subhalo number density profiles of the earliest forming haloes of any mass are quite similar to each other: they are parallel to the scaled mass density profile of their respective haloes until $r/R_h \sim 0.3$, where such haloes typically formed, and then rapidly increase reaching the μ profile of purely accreting haloes at a finite radius substantially smaller than R_h. On the contrary, the scaled profiles of the latest-forming haloes keep their initial trend parallel to the density profile of the respective haloes until a much larger radius, where they suddenly recover to reach the μ profile of purely accreting haloes at R_h. Thus, the two scaled subhalo number density profiles are quite distinct, meaning that this property would be a good tool for probing the halo formation time. However, since these profiles are scaled to the total number of subhaloes of each mass, they do not inform on the subhalo richness of haloes.

An alternative estimate of the radial distribution of subhaloes that is non-scaled and easier to determine in simulations as well as observations because using cumulative quantities is that put forward by Bose et al. (2020), namely the profile of the total number of subhaloes with masses below some given value that lies inside each
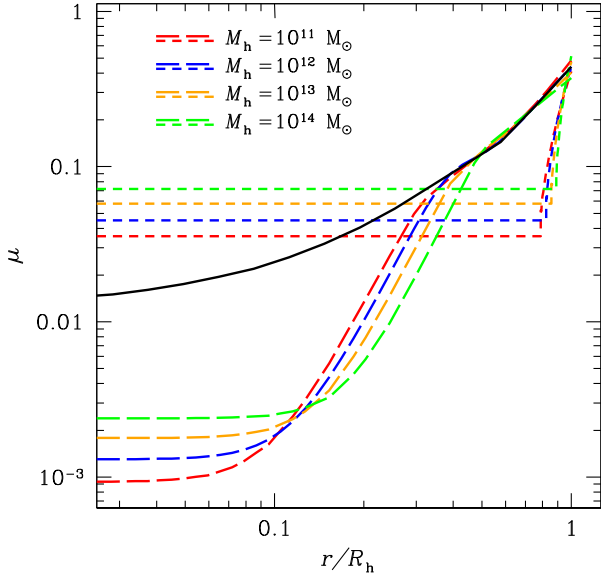
Figure 8. Mean truncated-to-original subhalo mass ratio profiles predicted using the CUSP M–c relation for the 20 per cent earliest forming haloes (long-dashed lines) and the 20 per cent latest-forming haloes (short-dashed lines) of several masses. For comparison, we plot the same profile for a purely accreting halo of 10^{12} M$_\odot$ (thick black line).
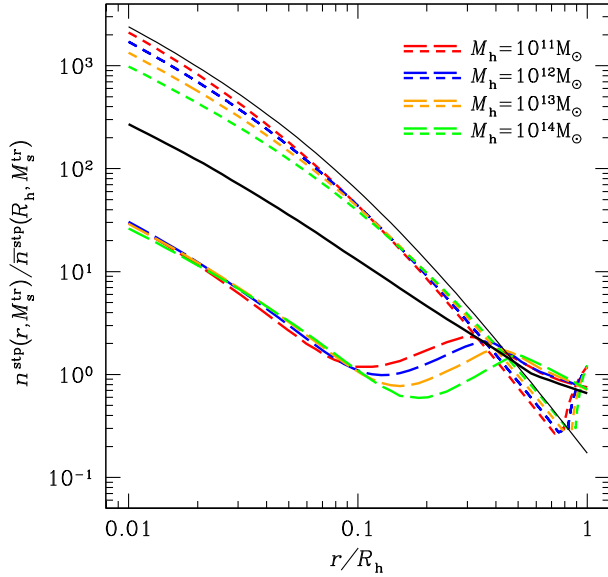(A colour version of this figure is available in the online journal.)

Figure 10. Integrated number density profiles of plain subhaloes with masses $10^{-6} M_h < M_s^{tr} < 10^{-3} M_h$ predicted by CUSP for the 20 per cent earliest forming (long-dashed lines) and 20 per cent latest-forming (short-dashed lines) haloes of different masses M_h multiplied by $[M_h/(10^{12}$ M$_\odot)]^{-0.08}$. For comparison, we plot the same profile for purely accreting haloes of the MW–mass (thick black line) and the integrated mass density profile of the (purely accreting or ordinary) halo (thin black line) with suited zero-point so as to essentially overlap with the profiles of the latest-forming haloes.
(A colour version of this figure is available in the online journal.)

Figure 9. Same as Fig. 8 (same lines) but for the scaled subhalo number density profiles. Also plotted is the formation time-independent mass density profile of haloes with 10^{12} M$_\odot$ (thin black line) scaled so as to facilitate the comparison.
(A colour version of this figure is available in the online journal.)

radius r. In Figs 10 and 11, we plot two versions of this integrated density profile: one dealing with 'plain' subhaloes as used so far, with masses in the range $10^{-6} M_h < M_s^{tr} < 10^{-3} M_h$ (the lower limit is supposed to account for the typical mass resolution of simulations of MW–mass haloes), and another one dealing with 'luminous' subhaloes, i.e. subhaloes harbouring faint and ultra-faint satellites (with stellar mass less than 3.2×10^6 M$_\odot$), ready to be compared
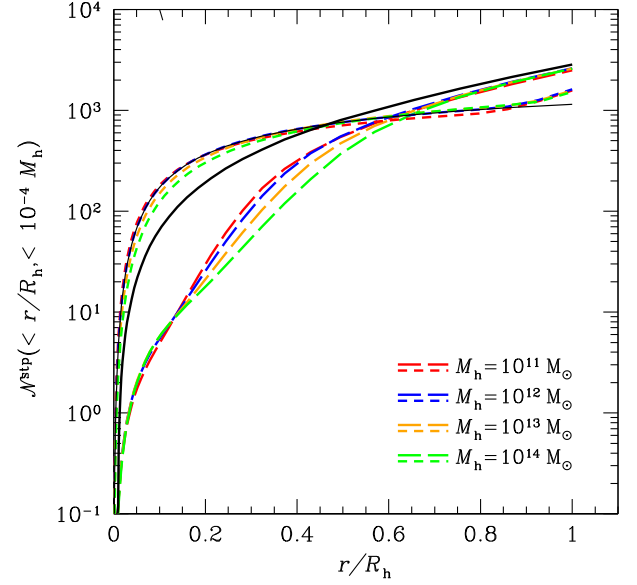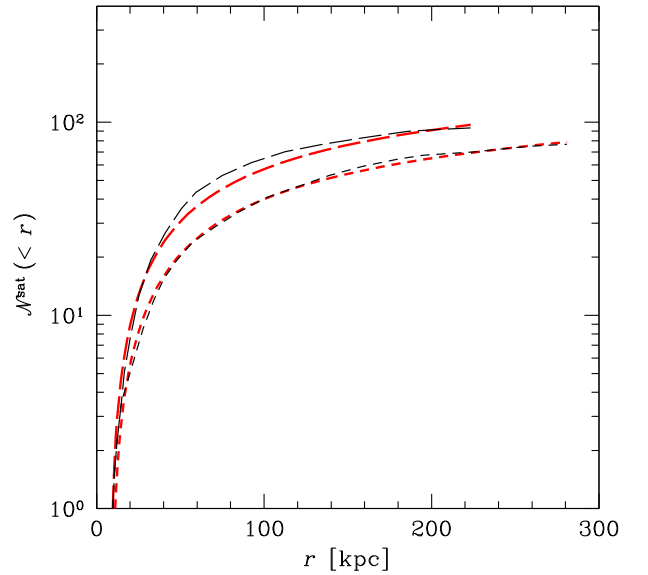

Figure 11. Integrated number density profiles of accreted subhaloes with masses 1.4×10^8 M$_\odot < M_s < 5 \times 10^8$ M$_\odot$ corresponding to satellites with stellar masses $M_\star < 10^6$ M$_\odot$ in haloes with $M_{200} = 1 \times 10^{12}$ M$_\odot$ (long-dashed red line) and $M_{200} = 1.3 \times 10^{12}$ M$_\odot$ (short-dashed red line) predicted by CUSP compared to the average profiles of the 20 per cent earliest forming (long-dashed black line) and 20 per cent latest-forming (short-dashed black line) haloes found by Bose et al. (2020) for MW–mass haloes ($M_{200} \sim 1 - 1.3 \times 10^{12}$ M$_\odot$).
(A colour version of this figure is available in the online journal.)

with the results of the simulations of MW–mass haloes carried by Bose et al. (2020) (see their fig. 5).

As can be seen in Fig. 10, the integrated radial number density profile of plain subhaloes in haloes of different masses essentially overlaps, in both the 20 per cent earliest and 20 per cent latest-forming haloes, when they are multiplied by $[M_{\rm h}/(10^{12}\ {\rm M_\odot})]^{-0.08}$; there is just a small dispersion at intermediate radii due to the fact that the formation time intervals cannot be scaled to $M_{\rm h}$. The fact that this mass dependence essentially coincides with that of the subhalo MFs[5] is not unsurprising: both properties arise from the integration over r of the original radial abundance of stripped subhaloes (equation [4]). Regarding the radial distributions in haloes of any fixed mass, we see that in the earliest forming haloes, it is slightly higher at large radii ($r/R_{\rm h} \gtrsim 0.6$) than in the latest-forming ones. Since subhaloes at those radii contribute the most to the total number of objects in the haloes, this means that the former haloes are richer than the latter ones (see the MFs of the two kinds of haloes below). But this is not the only difference: the profile of the earliest forming haloes decreases towards the halo centre much more steeply than that of the latest-forming haloes, which essentially stay parallel to the integrated mass density profile of the halo. As a consequence, both profiles cross at $r \sim 0.06R_{\rm h}$. Thus, the integrated radial distribution of plain subhaloes would be a very good probe of halo assembly history provided subhalo masses could be determined.

In Fig. 11, we compare the theoretical radial distributions of luminous subhaloes to those of satellites found by Bose et al. (2020) in the 20 per cent earliest and latest-forming MW–mass haloes ($M_{200} \sim 1 - 1.3 \times 10^{12}\ {\rm M_\odot}$). In principle, to calculate such theoretical profiles, we should find the upper- and lower-mass limits of stripped subhaloes that correspond to the upper and lower stellar mass limits of $3.2 \times 10^6\ {\rm M_\odot}$ and zero ${\rm M_\odot}$, respectively, used by Bose et al. But this correspondence depends on the radius in an unknown way, so that calculation is hard to achieve. Fortunately, there is an alternative way to calculate the desired radial distributions. The vast majority of satellites with stellar masses below $\sim 3.2 \times 10^6\ {\rm M_\odot}$ necessarily formed in haloes with masses between the minimum mass for star formation, $M_{\rm s} \sim 1.4 \times 10^8\ {\rm M_\odot}$ (Bose et al. 2020), and the maximum mass of haloes having formed $3.2 \times 10^6\ {\rm M_\odot}$ of stars, $\sim 5 \times 10^8$ ${\rm M_\odot}$. Certainly, a small fraction of satellites with a final stellar mass slightly below $3.2 \times 10^6\ {\rm M_\odot}$ may have suffered very strong tidal stripping affecting not only dark matter but also stars so that the initial mass of their subhaloes may be slightly larger than 5×10^8 ${\rm M_\odot}$. Consequently, the upper mass limit is actually somewhat fuzzy. None the less, since the radius of the stellar component is typically 1 per cent of the radius of their host haloes (Kravtsov 2013), the fraction of satellites with initial stellar mass larger than $\mu M_{\rm s}$ able to lose stars is very small. In addition, any small variation in the upper mass limit has an insignificant effect on the total number of haloes within the bracketed mass range. (Only do variations in the lower-mass limit substantially affect that number due to the much higher abundance of objects at the low-mass end.) Therefore, the number of accreted subhaloes within that mass range is a very robust estimate for the number of satellites with final masses below $3.2 \times 10^6\ {\rm M_\odot}$. And, since low-mass subhaloes, unaffected by dynamical friction, stay at the same apocentric radius as when they were accreted, the integrated radial distribution of satellites with masses in the included sample should essentially coincide with that of *accreted* subhaloes in the corresponding mass range independent of radius. We have

confirmed the robustness of this counting against small changes in the upper subhalo mass limit.

Certainly, the previous reasoning seems to contradict the results of simulations. As discussed in Paper I (see also Salvador-Solé et al. in preparation), the number density profile of accreted subhaloes of any mass is independent of the formation time of the host halo. Yet, Bose et al. (2020) found a significant difference between the radial distributions of satellites in the 20 per cent earliest and 20 per cent latest-forming haloes. The clue for that apparent contradiction is the different formation time PDF of haloes of different masses (Fig. 4). According to it, the 20 per cent earliest and 20 per cent latest-forming haloes should essentially coincide with the less and most massive objects, respectively, in the sample of haloes with masses $M_{200} \sim 1 - 1.3 \times 10^{12}\ {\rm M_\odot}$ used by Bose et al. (2020). And, since the abundance of accreted subhaloes within any given mass range is proportional to $M_{\rm h}^{-1}$ (see Section 2), the 20 per cent earliest forming haloes should be about 1.3 times richer than the 20 per cent latest-forming ones, where 1.3 is the ratio of extreme halo masses in the sample.

In Fig. 11, the integrated radial distributions of luminous subhaloes or, more exactly, the corresponding accreted subhaloes predicted by CUSP for haloes with the two extreme masses are compared to the integrated radial distributions of satellites (including orphan objects) found by Bose et al. (2020) for the 20 per cent earliest and 20 per cent latest-forming haloes with $M_{200} \sim 1 - 1.3 \times 10^{12}\ {\rm M_\odot}$. These authors provide the results obtained from the *Copernicus Complexio Low Resolution* (COLOR) simulation and the *Copernicus Complexio* (COCO) suite of simulations, with substantially higher resolution but much poorer statistics (COLOR has ~ 20 times more haloes than COCO). In any case, the mass resolution in both simulations is sufficient to detect star formation in all haloes with masses larger than $1.4 \times 10^8\ {\rm M_\odot}$, so including orphan satellites in stripped subhaloes with ending masses below the resolution mass should be sufficient to count all satellites. Yet, the higher resolution of the COCO simulations yields twice more satellites of the relevant masses than in the COLOR simulations (see fig. 1 of Bose et al. 2020). Since all-level subhaloes are also twice more abundant than first-level ones (Paper I), that result suggests that many satellites in the COCO simulations lie in high-level subhaloes. This would explain why including one orphan satellite per disrupted (first-level) subhalo in the COLOR simulation is not enough to recover the satellite abundance found in the COCO simulations. This conclusion is also supported by the fact that the satellite abundance in the COCO simulations agrees with the (all-level) subhalo abundance found in merger tree Monte Carlo simulations according to the EPS formalism (Bose et al. 2020). But the idea that 50 per cent of all satellites lie in high-level subhaloes is little realistic if we think about MW satellites. We have thus chosen to compare the radial distribution of (first-level) accreted subhaloes predicted by CUSP to the radial distribution of satellites found in the COLOR simulation.

As can be seen in Fig. 11, the predicted profiles almost fully agree, indeed, with those found by Bose et al. (2020). The only slight difference in the profiles of the earliest forming haloes is likely due to the effects of dynamical friction, excluded from our treatment. The upper subhalo mass limit of $\sim 5 \times 10^8\ {\rm M_\odot}$ is a factor of ~ 5 higher than the minimum mass of subhaloes suffering significant dynamical friction, so in early-forming haloes subhaloes could indeed be slightly more concentrated towards the centre. In any event, this agreement gives strong support to the conclusion that the radial distributions of satellites in haloes with masses in a very narrow range around any fixed value do not depend on their formation times. In addition, it reinforces the idea that satellites in the COLOR simulation lie

---

[5]It is again 0.08 rather than 0.07 due to the narrower formation time interval used here compared to that used in the subhalo MF of ordinary haloes.
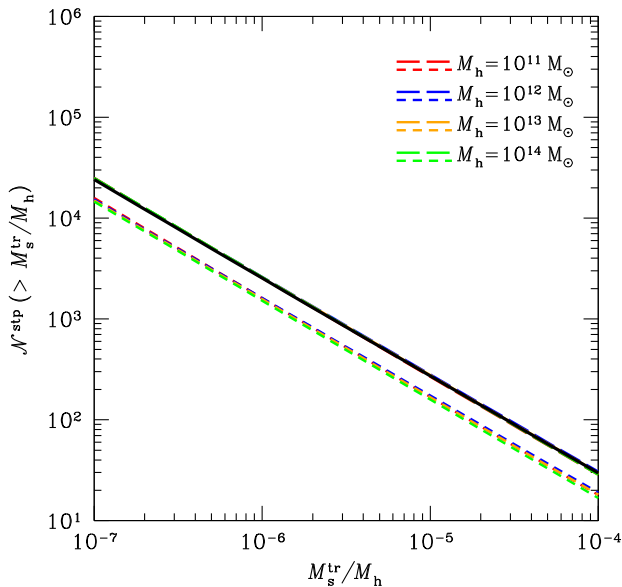
Figure 12. Same as Fig. 10 but for the cumulative subhalo MFs (also multiplied by $[M_h/(10^{12}\ \mathrm{M_\odot})]^{-0.08}$.
(A colour version of this figure is available in the online journal.)

in first-level subhaloes only, while those in the COCO simulations likely also occupy higher-level subhaloes.

The inclusion of orphan satellites in simulations is crucial for the previous comparison to be meaningful. Otherwise, stripping would affect the radial distribution of satellites through the varying number of disrupted subhaloes produced (e.g. Graus et al. 2019) and we could not use accreted subhaloes to count satellites. This does not mean, of course, that we can use the radial profiles of *non-orphan* satellites to probe halo formation times. Even though this is in principle possible in simulations, it is not in observations: reaching higher magnitudes allows a better determination of the radial distributions (and MFs) of satellites (e.g. Grand et al. 2021), but it does not alter the fundamental result that the abundance of satellites coincides with that of accreted subhaloes, independent of halo formation time. A better determination of the radial distribution of ultra-faint satellites should be useful, however, in connection with the 'missing satellite problem' itself. By comparing the radial distribution (or the MF) of ultra-faint satellites to the theoretical one(s) of the corresponding low-mass *accreted subhaloes* (equation 3, with $\mathcal{N}^{\mathrm{acc}}(M_s)$ given by the well-normalized MF of accreted subhaloes; see Paper I), it should be possible to confirm whether or not there is a problem in the satellite abundance. Note that such a test, unaffected by the complications arising from stripping and dynamical friction, has the added advantage that it is independent of the assembly history of haloes. To do that, it would be advisable, however, to exclude the very central region of haloes so as to avoid the effect of satellite destruction by central galaxies (see Bose et al. 2020 and references therein).

Finally, integrating the radial abundances, $\mathcal{N}^{t_f \pm \Delta}_{[M_h, t_h]}(r, M_s)$ out to R_h, we are led to the differential MF of plain or luminous subhaloes in haloes of different formation times in early-forming and late-forming haloes of any fixed mass. In Fig. 12, we depict the cumulative MFs of plain subhaloes for haloes of different masses. As usual, all MFs depend on halo mass as $M_h^{0.08}$. Regarding their similarity with the respective MF of purely accreting haloes, we see that, as expected, the MFs in the earliest forming haloes of any mass overlap with it,

while those of the latest-forming haloes are substantially lower. The difference is consistent with that found in the radial distribution of subhaloes at large radii, while the shape of all these MFs is always the same, essentially a power law of index close to -1. In this sense, subhalo MF is a poorer probe of halo formation time than the radial distribution of subhaloes.

5 SUMMARY AND DISCUSSION

The present paper is the last one of a series of three devoted to provide a comprehensive insight on halo substructure. Making use of the CUSP formalism, we derived, in Paper I, the properties of accreted subhaloes from the statistics of their seeds (nested peaks) in the primordial Gaussian random density field. In Paper II, we developed a detailed stripping model and used the properties of accreted subhaloes to derive those of stripped ones taking into account the repetitive tidal truncation and shock heating they suffer as they orbit within their hosts. As argued in Paper II, the properties of stripped subhaloes depend on the particular assembly history of their host halo. Nevertheless, their derivation was carried in the simplest, little realistic case of purely accreting haloes. Here, we have extended those results to ordinary haloes, i.e. having undergone major mergers, with the aim to: (1) see how the properties of substructure determined in Paper II are modified by the action of major mergers, (2) determine their dependence on substructure on halo mass and formation time (i.e. the time of the last major merger), and (3) find the capabilities of substructure as a probe of halo assembly history.

In a formal plane, it has been shown that all the properties of substructure in haloes of any mass and formation time are encoded in the mean truncated-to-original subhalo mass ratio profile, $\mu(r, M_s^{\mathrm{tr}})$, weakly dependent, actually, on M_s^{tr}. This profile is flat inside the radius reached by the object at its last major merger (inside which the system was completely scrambled) and rapidly approaches, at larger radii, the μ profile of the purely accreting halo of the same mass, setting the correspondence between r and t, which can be used to determine the time of the merger. This profile is at the base of all the remaining more practical results.

Regarding our first goal, we have demonstrated that the agreement found in Paper II between our predictions for purely accreting haloes and the results of simulations dealing with ordinary ones was not casual. Even though major mergers affect the properties of substructure in ordinary haloes with respect to purely accreting ones, their subhalo MF remains essentially unaltered. Major mergers leave only a substantial imprint in the radial distribution of subhaloes.

These results have important repercussions on our second goal. On the one hand, the dependence on halo mass of the subhalo MF is the same in ordinary haloes as in purely accreting ones. Since in the latter the formation time plays no role, that dependence can arise only from the mass dependence on mass of halo concentration, as found in Paper II. This dependence on halo mass is already seen, of course, in the (non-scaled) radial distribution of subhaloes whose integral over the radius leads to the MF. But the radial distribution of subhaloes in haloes of any given mass has the added interest that it also harbours information on the formation time of the host halo. Indeed, the scaled subhalo number density profile in ordinary haloes, independent of subhalo mass, is steeper than in purely accreting haloes, being proportional to the density profile of the halo inside the scrambling radius reached at its formation. This scaled version of the radial distribution of subhaloes does not inform on the subhalo richness, but another non-scaled version of it, the integrated subhalo number density profile, does. We have shown that this latter profile is higher at large radii contributing the most to

the total subhalo abundance in early-forming haloes than in late-forming ones. Moreover, the profiles in the two kinds of haloes are also very different as they cross each other at some intermediate radii. Of course, the different richness of haloes of a given mass is also reflected in their subhalo MF, which in extremely late-forming haloes is found to be substantially lower than in all the rest.

All these results lead to the following conclusion regarding the third goal: the radial distribution of subhaloes is very useful to probe the halo formation time, much better than the subhalo MF. Unfortunately, this conclusion only holds for plain subhaloes, not for satellites. Indeed, the properties of satellites do not depend on how their host halo stripped subhaloes. They depend only on the masses of accreted subhaloes where they formed, whose properties are independent of the halo formation time. Thus, the properties of ultra-faint satellites do not inform on the formation time of the host halo. The good news is that by simply comparing their radial distribution to that of very low-mass subhaloes they should unambiguously tell whether or not there really is a 'missing satellite problem', without depending on the particular formation history of the host halo.

In its current form, our analytic treatment does not account for dynamical friction. Of course, this is not a drawback when dealing with low enough mass subhaloes (faint enough satellites). But, if we want to deal with more massive subhaloes (more luminous satellites), we should account for that process. That possibility would be very welcome because, even though the radial distribution of satellites does not depend on subhalo stripping, it is sensitive to dynamical friction. Thus, the properties of satellites could still inform on the halo assembly history through the effect of dynamical friction. Work in this line is currently in progress (Salvador-Solé et al. in preparation).

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983
Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, ApJ, 304, 15
Bennet P., Sand D. J., Crnojević D., Spekkens K., Karunakaran A., Zaritsky D., Mutlu-Pakdil B., 2019, ApJ, 885, 153
Bennet P., Sand D. J., Crnojević D., Spekkens K., Karunakaran A., Zaritsky D., Mutlu-Pakdil B., 2020, ApJ, 893, L9
Benson A. J. et al., 2013, MNRAS, 428, 1774
Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, ApJ, 379, 440
Bose S., Hellwing W. A., Frenk C. S., Jenkins A., Lovell M. R., Helly J. C., Li B., 2016, MNRAS, 455, 318
Bose S., Deason A. J., Belokurov V., Frenk C. S., 2020, MNRAS, 495, 743
Bower R. G., 1991, MNRAS, 248, 332
Boylan-Kolchin M., Springer V., White S. D. M., Jenkins A., 2010, MNRAS, 406, 896
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bullock J. S., Boylan-Kolchin M., 2017, ARA&A, 55, 343
Carlsten S. G., Greene J. E., Peter A. H. G., Greco J. P., Beaton R. L., 2020, ApJ, 902, 124
Carlsten S. G., Greene J. E., Peter A. H. G., Beaton R. L., Greco J. P., 2021, ApJ, 908, 109
Cautun M., Hellwing W. A., van de Weygaert R., Frenk C. S., Jones B. J. T., Sawala W., 2014, MNRAS, 445, 2049
Crnojević D. et al., 2019, ApJ, 872, 80
Danieli S., van Dokkum P., Merritt A., Abraham R., Zhang J., Karachentsev I. D., Makarova L. N., 2017, ApJ, 837, 136
Deason A. J., Belokurov V., Evans N. W., Johnston K. V., 2013, ApJ, 763, 113
Diemand J., Kuhlen M., Madau P., 2007, ApJ, 657, 267
Elahi P. J., Widrow L. M., Thacker R. J., 2009, Ph. Rev. D, 80, 123513
Engler C. et al., 2021, MNRAS, 507, 4211
Fakhouri O., Ma C.-P., Boylan-Kolchin M., 2010, MNRAS, 406, 2267
Font A. S. et al., 2020, MNRAS, 498, 1765
Font A. S., McCarthy I. G., Belokurov V., 2021, MNRAS, 505, 783
Gaia Collaboration 2018, A&A, 616, A1
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
Gao L., Frenk C. S., Boylan-Kolchin M., Jenkins A., Springel V., White S. D. M., 2011, MNRAS, 410, 2309
Gao L., Frenk C. S., Jenkins A., Springel V., White S. D. M., 2012, MNRAS, 419, 1721
Geha M. et al., 2017, ApJ, 847, 4
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
Grand R. J. J. et al., 2021, MNRAS, 507, 4953
Graus A. S., Bullock J. S., Kelley T., Boylan-Kolchin M., Garrison-Kimmel S., Qi Y., 2019, MNRAS, 488, 4585
Green S. B., van den Bosch F. C., 2019, MNRAS, 490, 2091
Griffen B. F., Ji A. P., Dooley G. A., Gómez F. A., Vogelsberger M., O'Shea B. W., Frebel A., 2016, ApJ, 818, 10
Han J., Cole S., Frenk C. S., Jing Y., 2016, MNRAS, 457, 1208
Han J., Cole S., Frenk C. S., Benitez-Llambay A., Helly J., 2018, MNRAS, 474, 604
Hellwing W. A., Frenk C. S., Cautun M., Bose S., Helly J., Jenkins A., Sawala T., Cytowski M., 2016, MNRAS, 457, 3492
Henry J. P., 2000, ApJ, 534, 565
Ishiyama T. et al., 2020, MNRAS, 506, 4210
Jiang F., van den Bosch F. C., 2016, MNRAS, 458, 2848
Jiang F., Dekel A., Freundlich J., van den Bosch F. C., Green S. B., Hopkins P. F., Benson A., Du X., 2021, MNRAS, 502, 621
Juan E., Salvador-Solé E., Domènec G., Manrique A., 2014, MNRAS, 439, 719
Juan E., Salvador-Solé E., Domènech G., Manrique A., 2014, MNRAS, 439, 3156
Kampakoglou M., Benson A. J., 2007, MNRAS, 374, 775
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Klypin A. A., Trujillo-Gomez S., Primack J., 2011, ApJ, 740, 102
Komatsu E. et al., 2011, ApJS, 192, 18
Kravtsov A. V., 2013, ApJ, 764, L31
Lacey C., Cole S., 1994, MNRAS, 271, 671
Lancaster L., Belokurov V., Evans N. W., 2019, MNRAS, 484, 2556
Lee J., 2004, ApJ, 604, L73
Lovell M. R., Frenk C. S., Eke V. R., Jenkins A., Gao L., TheunsT., 2014, MNRAS, 439, 300
Lovell M. R., Cautun M., Frenk C. S., Hellwing W. A., Newton O., 2021, MNRAS, 507, 4826
Manrique A., Salvador-Solé E., 1995, ApJ, 453, 6
Manrique A., Salvador-Solé E., 1996, ApJ, 467, 504
Manrique A., Raig A., Solanes J. M., González-Casado G., Stein P., Salvador-Solé E., 1998, ApJ, 499, 548
Mao Y.-Y., Geha M., Wechsler R. H., Weiner B., Tollerud E. J., Nadler E. O., Kallivayalil N., 2020, ApJ, 907, 85

Moore B., Ghigna S., Governato F., Lake G., Quinn T., Stadel J., Tozzi P., 1999, ApJ, 524, L19

Navarro J. F., Frenk C. S., White S. D. M., 1997, ApJ, 490, 493

Oguri M., Lee J., 2004, MNRAS, 355, 120

Onions J. et al., 2012, MNRAS, 423, 1200

Peñarrubia J., Benson A. J., 2005, MNRAS, 364, 977

Press W. H., Schechter P., 1974, ApJ, 187, 425

Pullen A. R., Benson A. J., Moustakas L. A., 2014, ApJ, 792, 24

Raig A., González-Casado G., Salvador-Solé E., 2001, MNRAS, 327, 939

Richings J. et al., 2020, MNRAS, 492, 5780

Rodríguez-Puebla A., Behroozi P., Primack J., Klypin A., Lee C., Hellinger D., 2016, MNRAS, 462, 893

Ruchti G. R. et al., 2015, MNRAS, 450, 2874

Salvador-Solé E., Manrique A., 2021, ApJ, 914, 141

Salvador-Solé E., Viñas J., Manrique A., Serra S., 2012a, MNRAS, 423, 2190

Salvador-Solé E., Serra S., Manrique A., González-Casado G., 2012b, MNRAS, 424, 3129

Salvador-Solé E., Manrique A., Botella I., 2021a, MNRAS, 509, 5305

Salvador-Solé E., Manrique A., Botella I., 2021b, MNRAS, 509, 5316

Samuel J. et al., 2020, MNRAS, 491, 1471

Sheth R. K., 2003, MNRAS, 345, 1200

Smercina A., Bell E. F., Price P. A., D'Souza R., Slater C. T., Bailin J., Monachesi A., Nidever D., 2018, ApJ, 863, 152

Springel V. et al., 2008a, MNRAS, 391, 1685

Sugiyama N., 1995, ApJS, 100, 281

Taylor J. E., Babul A., 2001, ApJ, 559, 716

Taylor J. E., Babul A., 2004, MNRAS, 348, 811

Tollerud E. J., Boylan-Kolchin M., Bullock J. S., 2014, MNRAS, 440, 3511

van den Bosch F. C., Jiang F., 2016, MNRAS, 458, 2870

van den Bosch F. C., Tormen G., Giocoli C., 2005, MNRAS, 359, 1029

van den Bosch F. C., Ogiya G., Hahn O., Burkert A., 2018, MNRAS, 474, 3043

Wang J. et al., 2011, MNRAS, 413, 1373

Wyse R. F. G., 2001, ASPC, 230, 71

Zentner A. R., Bullock J. S., 2003, ApJ, 598, 49

Zentner A. R., Berlin A. A., Bullock J. S., Kravtsov A. V., Wechsler R. H., 2005, ApJ, 624, 505

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