On stellar mass loss from galaxies in groups and clusters

Édouard Tollet,[1,2]★ Andrea Cattaneo,[1,3] Gary A. Mamon,[3] Thibaud Moutard[4] and Frank C. van den Bosch[5]

[1]*Observatoire de Paris, GEPI, CNRS and PSL Research University, 61 Avenue de l'observatoire, F-75014 Paris, France*

[2]*Université Paris Diderot, Sorbonne Paris Cité, F-75013 Paris, France*

[3]*Institut d'Astrophysique de Paris, CNRS and UPMC, 98 bis Boulevard Arago, F-75014 Paris, France*

[4]*Department of Astronomy and Physics, Saint Mary's University, 923 Robie Street, Halifax, Nova Scotia B3H 3C3, Canada*

[5]*Astronomy Department, Yale University, PO Box 208101, New Haven, CT 06520-8101, USA*

ABSTRACT

We estimate the stellar mass that satellite galaxies lose once they enter groups (and clusters) by identifying groups in a high-resolution cosmological *N*-body simulation, assigning entry masses to satellite galaxies with halo abundance matching at the entry time and comparing the predicted conditional stellar mass function of satellite galaxies at $z \simeq 0$ with observations. Our results depend on the mass of the stars that form in satellite galaxies after the entry time. A model in which star formation shuts down completely as soon a galaxy enters a group environment is ruled out because it underpredicts the stellar masses of satellite galaxies even in the absence of tidal stripping. The greater is the stellar mass that can form, the greater the fraction that needs to be tidally stripped. The stellar mass fraction lost by satellite galaxies after entering a group or cluster environment is consistent with any value in the range 0–40 per cent. To place stronger constraints, we consider a more refined model of tidal stripping of galaxies on elongated orbits (where stripping occurs at orbit pericentres). Our model predicts less tidal stripping: satellite galaxies lose ∼20–25 per cent of their stellar mass since their entry into the group. This finding is consistent with a slow-starvation delayed-quenching picture, in which galaxies that enter a group or cluster environment keep forming stars until at least the first pericentric passage.

Key words: galaxies: evolution – galaxies: formation – galaxies: interactions.

1 INTRODUCTION

1.1 Constraints from the intracluster light

The notion of intracluster light (ICL) began to emerge after Zwicky (1957) and Welch & Sastry (1971) detected a diffuse luminous background around NCG 4874 and NGC 4889, the two supergiant elliptical galaxies that dominate the central region of the Coma cluster. Since central dominant (cD) galaxies with extensive outer envelopes are a common occurrence in very rich clusters (Matthews, Morgan & Schmidt 1964; Morgan & Lesh 1965; Bautz & Morgan 1970), Gallagher & Ostriker (1972) interpreted the ICL at the centre of Coma as 'a diffuse intergalactic cloud of stars evaporated from colliding galaxies', which 'will contribute to the formation of a cD system's envelope or in fact may constitute the cD "galaxy" itself'. Analytical calculations and *N*-body simulations have confirmed that tidal stripping by companion galaxies and by the cluster's potential can explain the origin of the ICL (e.g. Merritt 1983; Mamon 1987;

Ghigna et al. 1998; Hayashi et al. 2003; Willman et al. 2004; Mihos et al. 2005; Abadi, Navarro & Steinmetz 2006; Purcell, Bullock & Zentner 2007). Klimentowski et al. (2009), Łokas, Kazantzidis & Mayer (2011) and Kazantzidis et al. (2011) found a median stellar mass loss of ∼30–35 per cent at each pericentric passage.

Forty years after Gallagher & Ostriker (1972), there is still no consensus whether the outer envelopes of cD galaxies belong to the ICL or to the galaxies themselves. The issue could be dismissed as largely semantic, but is also the reason why the galaxy stellar mass functions (SMFs) of Baldry et al. (2012) and Bernardi et al. (2013) differ by ∼0.5 dex at high masses (but see Bernardi et al. 2017).

The ICL contributes ∼10–30 per cent of a cluster's total luminosity (Gonzalez, Zabludoff & Zaritsky 2005; Zibetti et al. 2005; Krick, Bernstein & Pimbblet 2006) and sometimes even more (Lin & Mohr 2004). However, stellar haloes formed out of disrupted satellites are also present in lower mass systems. In particular, it has become clear that the stellar halo of the Milky Way contains considerable substructure in the form of stellar streams (Helmi et al. 1999; Yanny et al. 2003; Bell et al. 2008). In some cases, the streams can be unambiguously associated with the satellite galaxies from which they came (Ibata, Gilmore & Irwin 1994; Odenkirchen et al. 2002).

*E-mail: edouard.tollet@obspm.fr

Similar streams have also been detected in our neighbour galaxy M31 (e.g. Ferguson et al. 2002). In this article, we assess the impact of tidal stripping on the stellar masses of galaxies not just in clusters but across a broad range of environments.

1.2 Constraints from the growth of cD galaxies

Conroy, Wechsler & Kravtsov (2007) studied the mass growth of cD galaxies from $z \sim 1$ to 0 by using the abundance matching (AM) technique (Marinoni & Hudson 2002; Vale & Ostriker 2004), described in detail by Behroozi, Conroy & Wechsler (2010), which assumes that halo masses (or equivalent properties) are strongly correlated with observable galaxy properties such as luminosity or stellar mass.

Let $n_*(m_*)$ be the galaxy SMF, which can be determined from observations, and let $n_h(M_h)$ be the halo mass function (HMF), which depends on the cosmology and can be determined either analytically (e.g. Press & Schechter 1974) or from N-body simulations. If the stellar mass m_* is a growing function of the halo mass M_h with negligible scatter,[1] then the number density of galaxies with stellar mass $> m_*$ will be equal to the number density of haloes with mass $> M_h$:

$$\int_{m_*}^{\infty} n_*(m'_*) \, \mathrm{d}m'_* = \int_{M_h}^{\infty} n_h(M'_h) \, \mathrm{d}M'_h. \tag{1}$$

Given the stellar masses (from observations), AM determines the halo masses by solving equation (1) for M_h, and conversely, given the halo masses (from a simulation), AM provides the stellar masses by solving equation (1) for m_*. In other words, the most massive halo is assigned the largest stellar mass, the second most massive halo is assigned the second largest stellar mass, and so on. The power of AM was nicely illustrated by Conroy, Wechsler & Kravtsov (2006), who painted galaxy luminosities on halo maximum circular velocities to predict very well the two-point correlation function of galaxies in bins of luminosity.

Conroy et al. (2007) used the results of AM at $z = 1$ to populate haloes at $z = 1$ with galaxies, and they followed these galaxies until $z = 0$. They asked what happens when a smaller halo disappears into a larger one due to hierarchical merging. Three possibilities were considered: (i) the galaxy in the smaller halo merges with the central galaxy of the larger halo; (ii) the galaxy in the smaller halo becomes a satellite galaxy in the larger halo; (iii) the galaxy in the smaller halo is disrupted and its stars become part of the ICL of the larger halo. The first assumption leads to cD galaxies those were far too bright. The second assumption underestimated the total luminosity of the central galaxy plus the ICL by about a magnitude. The third assumption was found to be in reasonably good agreement with the observations.

Kang & van den Bosch (2008) and Cattaneo et al. (2011) provided independent arguments in support of Conroy et al.'s (2007) conclusion. Kang & van den Bosch (2008) argued that tidal disruption is necessary to avoid that mergers with bluer satellites spoil the colours of massive red galaxies. In Cattaneo et al. (2011), we used a method intermediate between semi-analytic and halo occupation distribution (HOD) modelling to quantify the relative importance of gas accretion and mergers in the mass growth of galaxies. We calibrated our model to fit the galaxy SMF around and below the

---

[1] Throughout this article, we use lower case letters for stellar masses and radii within subhaloes, and upper case letters for dark matter (DM) masses and subhalo radial coordinates within haloes.

knee of the Schechter function, where most of the galaxies are, and we found that the SMF above a stellar mass of $3 \times 10^{11} \, \mathrm{M}_\odot$ was overestimated by ~ 0.2 dex. This discrepancy disappeared when we included a simple model of tidal stripping, assuming a fixed relative loss of stellar mass at each orbit, calibrated on the simulations of Klimentowski et al. (2009).

1.3 Constraints from satellite galaxies

While Conroy et al. (2007) and Cattaneo et al. (2011) focused on tides as a mechanism to prevent overgrowth of cD galaxies between $z \sim 1$ and 0, Liu et al. (2010) compared the predictions of three semi-analytical models (SAMs, Kang et al. 2005; Bower et al. 2006; De Lucia & Blaizot 2007) with the conditional SMF of satellite galaxies in Sloan Digital Sky Survey (SDSS) groups (the conditional SMF $\phi(m_*|M_h)$ is defined so that $\mathrm{d}\phi$ is the number of satellite galaxies with stellar mass between m_* and $m_* + \mathrm{d}m_*$ in a host system of halo mass M_h). In all three models, the number of satellite galaxies was systematically overpredicted, particularly at low halo masses. Tidal stripping was considered as a possible explanation. A mechanism to reduce the number of satellites in massive haloes is also necessary to bring SAMs in agreement with the observed clustering properties of red galaxies (de la Torre et al. 2011).

The galaxy stellar mass–metallicity relation is another source of observational evidence. Galaxies that are more massive have higher metal abundances (e.g. Gallazzi et al. 2005). Pasquali et al. (2010) found that satellite galaxies have higher metallicity than central galaxies of the same mass. They interpreted this observation as a consequence of tidal stripping, which has reduced the stellar masses of satellite galaxies, while preserving their stellar metallicities. Henriques & Thomas (2010) confirmed that incorporating tidal disruption improves the agreement with the mass–metallicity relation, but their interpretation of this finding differs from that of Pasquali et al. (2010). For Pasquali et al. (2010), tidal stripping of stars changes the position of galaxies on the stellar mass–metallicity relation by reducing their masses at constant metallicity. However, ram-pressure and tidal stripping of gas can produce a similar shift by increasing metallicity at constant stellar mass (starvation of gas accretion shuts down star formation and causes galaxies to behave like close boxes; Peng, Maiolino & Cochrane 2015). Collectively, these articles highlight the difficulty of disentangling the effects of stripping and starvation.

1.4 This work

In this article, we build on Conroy et al.'s (2007) method and extend its scope to the investigation of the conditional SMF of satellites. We start by identifying group and cluster haloes in a cosmological N-body simulation of dissipationless hierarchical clustering. Merger trees extracted from the simulation allow us to follow galaxies from the entry time to the present. We use AM to assign stellar masses to satellite galaxies at the entry time. By comparing the distribution of entry stellar masses m_{entry} to the distribution of satellite masses m_s observed in the Universe today (Yang, Mo & van den Bosch 2009; Yang et al. 2012), we can derive a *lower* limit $m_{\mathrm{entry}} - m_s$ to the stellar mass lost through tidal stripping (the lower limit is zero if statistically $m_{\mathrm{entry}} \lesssim m_s$).

For a better estimate of the stellar mass $\Delta m_{\mathrm{strip}}$ lost through tidal stripping, we must increment this lower limit by the mass Δm_* of the stars formed in the satellite after the entry time:

$$\Delta m_{\mathrm{strip}} = m_{\mathrm{entry}} + \Delta m_* - m_s. \tag{2}$$

While $m_{\rm entry}$ is determined from AM at the entry time, estimating Δm_* requires additional assumptions. We can however derive an *upper* limit for Δm_* by assuming that the masses of satellite galaxies grow with those of their subhaloes following the same stellar mass–halo mass (SMHM) relation established for central galaxies from AM (see Section 5.1 for more details). We can therefore determine a lower and an upper limit for $\Delta m_{\rm strip}$, which we can compare to a theoretical estimate of $\Delta m_{\rm strip}$ based on a model described in Section 5.2.

Implementing this research programme requires an accurate determination of the sub-HMF and of the orbits of subhaloes in groups and clusters (the strength of the tides depends on the pericentric radius). In this article, we describe in detail how we solve the technical problem of reconstructing the orbits of *orphan galaxies*, the subhaloes of which are no longer resolved by the *N*-body simulation. We have tested the convergence of our scheme by comparing our results when we use merger trees from a simulation with 512^3 particles and from another with 1024^3 particles, both of which were run for the same cosmology and the same initial conditions.

The plan of the article is thus as follows. In Section 2, we describe the *N*-body simulation and the way we analyse it (identification of haloes and subhaloes, measurement of halo properties and construction of merger trees). In Section 3, we present our scheme to handle orphan galaxies (how we compute their orbits and how we decide at which time they merge with the central galaxy). In Section 4, we explain how we use AM to compute $m_*(M_{\rm h}, z)$, the stellar mass of the central galaxy in a halo of mass $M_{\rm h}$ at redshift z. In Section 5, we elaborate on our two different models to assign stellar masses to subhaloes: one in which $m_{\rm s}$ grows following the same relation for central galaxies, the other in which there is a complete shutdown of star formation in groups and clusters. We also present our model (following the simpler model of Mamon 2000) for computing the stellar mass stripped from galaxies in a simple form of the impulsive stripping approximation, where stripping occurs instantaneously at the pericentric passage. In Section 6, we compare our predicted distribution for $m_{\rm entry} + \Delta m_*$ and $m_{\rm entry} + \Delta m_* - \Delta m_{\rm strip}$ with observations of the conditional SMF (the distribution for $m_{\rm s}$ as a function of the global environment). For each model, we also compare the results of our calculations with the masses of central galaxies. Finally, in Section 7, we discuss the uncertainties that affect our results and summarize the conclusions of the article.

2 N-BODY SIMULATION AND MERGER-TREE EXTRACTION

We use a cosmological *N*-body simulation with $\Omega_{\rm m} = 0.308$, $\Omega_\Lambda = 0.692$, $\Omega_{\rm b} = 0.0481$, $\sigma_8 = 0.807$ and $H_0 = 67.8 \, {\rm km \, s^{-1} \, Mpc^{-1}}$ (Planck Collaboration XX 2014, *Planck* + WP + BAO). The simulation has a computational volume of $(100 \, {\rm Mpc})^3$ and contains 1024^3 particles. The same simulation, with the same initial conditions, has also been run with 512^3 particles to test for convergence.

287 snapshots (regularly spaced in the logarithm of the expansion factor) were saved to disc from $z = 16.7$ to 0. The corresponding output times are in steps of 145 Myr (at $z = 0$) or smaller. We have processed each snapshot with the halo finder HALOMAKER (Tweed et al. 2009), which is based on ADAPTAHOP (Aubert, Pichon & Colombi 2004). ADAPTAHOP is an excursion and percolation algorithm. It selects all particles above a density threshold and links each one to its 32 nearest neighbours. If the density distribution within a halo has more than one peak separated by saddle points, ADAPTAHOP decomposes it into a main host halo and a hierarchy of subhaloes, sub-subhaloes, etc. For simplicity of language, we shall

refer to all substructures as subhaloes independently of their rank in the hierarchy. The halo masses that we measure from the *N*-body simulation are exclusive, i.e. they do not include those of subhaloes. By construction, a host halo is always more massive than its most massive subhalo.

We further assume that, to belong to a halo or subhalo, a particle must be gravitationally bound to it.

For each halo containing at least 100 bound particles, we determine the inertia ellipsoid, which is centred on its centre of mass, and we rescale it until the overdensity, defined as the mean density inside the inertia ellipsoid divided by the critical density of the Universe, equals overdensity contrast given by the fitting formulae of Bryan & Norman (1998) in the case a *Planck* cosmology ($\Delta_c = 102$ at $z = 0$).[2] The halo virial mass $M_{\rm h}$ is the mass of the gravitationally bound *N*-body particles contained within the virial ellipsoid (i.e. the rescaled inertia ellipsoid). The virial radius $R_{\rm vir} = (a\,b\,c)^{1/3}$ is that of a sphere whose volume equals that of the virial ellipsoid of semi-axes a, b and c. We fit the spherically averaged density distribution of each halo with the NFW profile (Navarro, Frenk & White 1996) to measure its concentration c.

In case of a subhalo, we use the same procedure to obtain a first estimate of its mass and radius. Then, we shrink the subhalo by peeling off its outer layers until the density at the recomputed radius $R_{\rm t}$ is at least as large as the host density at the position where the subhalo is located. The subhalo mass $M_{\rm s}$ and the concentration parameter c is recomputed accordingly. The particles peeled off the outer layers of a subhalo are re-assigned to the host halo if they are gravitationally bound to it.

The TREEMAKER algorithm (Tweed et al. 2009) is used to link haloes/subhaloes identified at different redshifts and generate merger trees. A halo is identified as the descendent of another when it inherits more than half of its progenitor's particles. Because of this definition, a halo can have many progenitors, but at most one descendent. The main progenitor is always the one with the largest virial mass. A halo/subhalo is found to have no descendent if it loses more than half its mass, but no single halo accretes enough mass from it to qualify as its descendent. Typically, this happens when a smaller halo crosses a larger one at high speed. In these cases, the subhalo may be no longer identified but its particles are not assigned to the larger one because they are not gravitationally bound to it. However, subhaloes that disappear without leaving any descendents are rare and of scarce statistical significance. Most of those subhaloes disappear close to the pericentre, where the contrast against the host is weakest, and a fraction of them is detected again by the halo finder after their passage. As these objects are a possible source of artefacts for our model, we decide to ignore all subhaloes that have never been detected as central or field halo in any previous snapshot.

3 ORPHAN GALAXIES AND GHOST SUBHALOES

When a halo enters a group or a cluster and becomes a subhalo, it begins to lose mass owing to the tides exerted by the gravitational potential of the host (Fig. 1). Eventually, the mass loss may be so large that the subhalo is no longer identified by the halo finder. In SAMs, galaxies associated with subhaloes that are no longer

---

[2] The formulae of Bryan & Norman (1998) are a fit to the predictions of the spherical top-hat collapse model, but we have assumed that small deviations from sphericity do not change the virial density of DM haloes.
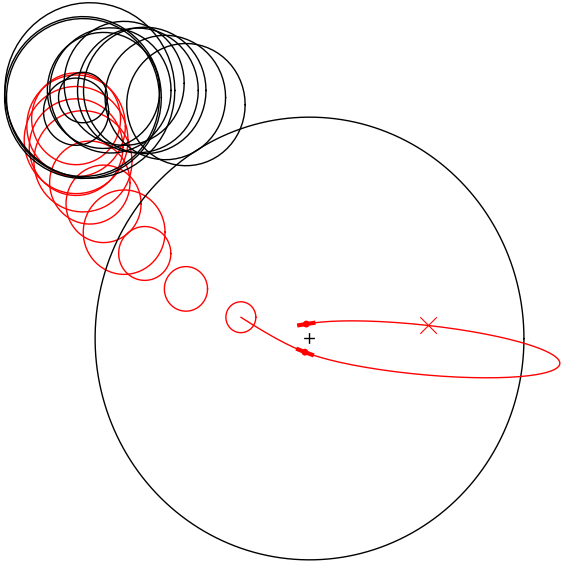
Figure 1. Trajectory and size of a small halo that becomes a subhalo of a larger one and eventually merges with it. Before the halo finder identifies the small halo as a subhalo, its virial radii are shown as *small black circles*. Their overlapping demonstrates how good the time resolution of our merger trees is. Once the halo finder identifies it as a subhalo (this occurs several time-steps before the subhalo enters the virial radius of the host halo, shown by the *large black circle*, the sizes of the tidal radii are shown as *red circles*, the smallest of which denotes the subhalo's last detection in the *N*-body simulation. The subhalo then becomes a ghost subhalo and its orbit (*solid red line*) is followed analytically, solving equation (4) in conjunction with equations (5) and (3). The two *small red filled circles* correspond to the first and the second pericentric passage of the ghost subhalo since the time of last detection. The *red cross* indicates the position of the subhalo when the dynamical-friction countdown timer comes to zero. The *black plus sign* denotes the centre of mass of the host system. The *thick* part of the solid line shows the part of the orbit around the pericentre along which the tides are supposed to act on the stars in the impulsive approximation.

identified are called *orphans* (Springel, Yoshida & White 2001; Guo et al. 2010). In reality, a subhalo still exists. We have simply lost our capacity to detect it. We therefore call it a *ghost* subhalo. Orphan galaxies may also be created by non-physical artefacts from the halo finder itself (Knebe et al. 2011; Srisawat et al. 2013; Avila et al. 2014).

Although our model is not a SAM, we face the same problem of deciding how long a galaxy will survive after it becomes an orphan. Immediately merging the orphan galaxy with the central galaxy of the host system may produce too few satellites and too massive central galaxies. The brute force solution is to increase the number of particles until the results converge above a specified mass. The most commonly followed alternative is to estimate the survival time analytically from the orbital decay time through dynamical friction, t_{df} (see Knebe et al. 2015; Pujol, Skibba & Gaztañaga 2017, for an overview of the orphan problem in SAMs of galaxy formation).

In its simplest version, this assumption is coupled to that of progressive decay on circular orbits (Somerville & Primack 1999; Hatton et al. 2003; Cattaneo et al. 2006; Cora 2006; Gargiulo et al. 2015). However, this approach neglects the high typical orbital elongations of satellites (Ghigna et al. 1998), which are important for our analysis because the strength of the tides depends on the pericentric radius. A more sophisticated approach is to track the defunct

subhalo's most bound particle until a time t_{df} has elapsed (De Lucia & Blaizot 2007; Benson 2012; Gonzalez-Perez et al. 2014). Here, we follow a third approach, first applied by Lee & Yi (2013), which consists of following the orbits of ghost suhaloes semi-analytically by integrating their equations of motion in presence of two forces: the gravitational attraction of the host halo (computed assuming an NFW profile) and dynamical friction (the physical reason why satellites lose energy, spiral in and eventually fall on to the central galaxy).

When a subhalo ceases to be identified by the halo finder (more precisely, when it is not the main progenitor of its descendant in the merger tree[3]), we save its mass M_s, position $\boldsymbol{R}$ and velocity $\boldsymbol{V}$ in the host halo's reference frame at the time of last detection. If the system that disappears is a halo, we treat its descendant's main progenitor as if it was its host. These values are the initial conditions from which we start integrating the equations of motions for the ghost subhalo (the red curve in Fig. 1 shows the orbit of a ghost subhalo computed in this manner after the subhalo was no longer resolved[4]).

We model the ghost subhalo as point particle of mass M_s that moves under the action of two forces: the gravitational attraction of the host and the dynamical-friction drag. However, we shall soon see that the dynamical-friction drag depends on M_s. Therefore, we cannot integrate the equations of motions without computing the evolution of M_s due to tidal stripping at each time-step. The structure of our calculation is thus as follows. In Section 3.1, we describe our method for computing tidal stripping of ghost haloes, which is based on the instantaneous-tide approximation (the tidal radius of the ghost subhalo at a given time is entirely determined by the configuration of the subhalo–host system at that time, although we never allow the tidal radius to grow again after a ghost halo has been stripped). In Section 3.2, we present the equations of motions that we integrate to compute the orbits of ghost subhaloes. Finally, in Section 3.3, we discuss the time at which we should stop their integration because the satellite galaxy associated with the ghost subhalo has merged with the central one.

3.1 Tidal stripping of ghost subhaloes

Once the halo finder identifies a structure as a subhalo, the radius it returns is no longer the virial radius but the tidal radius R_t, computed with the equation:

$$\frac{M_s(R_t)}{R_t^3} = |\alpha| \frac{M_h(R)}{R^3}, \tag{3}$$

where M_s is the subhalo mass within R_t, R is the distance of the subhalo from the centre of mass of the host, $M_h(R)$ is the host-halo mass within R and $|\alpha| = 1$.

---

[3] The merger tree is constructed by linking haloes that the halo finder has identified in output files at different time-steps. A halo can have more than one progenitor but at most one descendent. If the halo fragments, its descendent is the fragment that has inherited more than half of its particle. If none of the fragments contain more than half of the particles that the halo finder had assigned to the halo at the previous time-step, the halo is considered to have disappeared. The halo or subhalo is also considered to have disappeared if tidal stripping has been so strong that it has lost more than half of its particles from one time-step to the next because the algorithm that constructs the tree is not able to recognize that the stripped halo is the descendent of its progenitor.

[4] The rapid decrease in size when passing from the black circles to the red circles in Fig. 1 is an artefact of the halo finder.

Equation (3) has a theoretical justification because it is the prediction of tidal theory in the approximation of circular orbits and instantaneous tide (Appendix A). However, according to this theory, α should be the local logarithmic slope of the DM mean-density profile ($\alpha < 0$ because density decreases with radius). For an NFW profile, $-2.6 \lesssim \alpha \lesssim -2.2$ for $R = R_{\rm vir}$, but numerical experiments in Appendix A suggest that the appropriate $|\alpha|$ (the one that gives the correct value of $R_{\rm t}$) is even higher ($\alpha \lesssim -3$). For $R \to 0, \alpha \to -1$ for an NFW profile. However, the presence of a massive central galaxy could imply that $\alpha \ll -1$ even at relatively small radii (Fig. A2). Therefore, the value $|\alpha| = 1$ assumed by the halo finder is therefore likely to overestimate $R_{\rm t}$, at least within the circular-orbit approximation. This is not a problem for the orbits of detected subhaloes, which are computed self-consistently by the *N*-body simulation, but it is a point that we must consider when computing the tidal radii of ghost subhaloes.

In this article, we compute $R_{\rm t}$ using $\alpha = -3$ for all ghost subhaloes and we never allow its value to grow again (although equation 3 is instantaneous and thus gives growing values of $R_{\rm t}$ between the pericentre and the apocentre). The implications of assuming $\alpha = -3$ will be discussed in Appendix A, after we have explained all the elements that enter our analysis.

3.2 Orbital motion

A ghost subhalo is assumed to move in the gravitational potential $\Phi(\mathbf{r})$ of the halo directly above it in the hierarchy of substructures. 10 per cent of ghosts are sub-subhaloes. For these systems, Φ is the gravitational potential of the subhalo that contains them. In 46 per cent of these cases (which correspond to 4.6 per cent of all ghost systems), the subhalo merges with its host before the ghost sub-subhalo merges with the subhalo. When this happens, the ghost sub-subhalo is promoted to ghost subhalo and continues its orbital motion in the gravitational potential of the host halo.

The equation of motion for a ghost subhalo is:

$$\dot{V} = -\nabla\Phi + \mathbf{a}_{\rm df},\tag{4}$$

where Φ is computed assuming an NFW profile for the density distribution $\rho_{\rm h}$ of the system directly above the subhalo in the hierarchy of substructures (heretofore, the host halo, even if it is a subhalo) and where

$$\mathbf{a}_{\rm df} = -\frac{4\pi\, G^2 \rho_{\rm h}\, M_{\rm s}\, \ln\Lambda}{V^3}\, f\left(\frac{V}{\sqrt{2}\,\sigma}\right)\, V\tag{5}$$

with

$$f(x) = {\rm erf}(x) - \frac{2}{\sqrt{\pi}}\, x\, \exp(-x^2)\tag{6}$$

is the acceleration due to the dynamical-friction force (Chandrasekhar 1943). In equation (5), $M_{\rm s}$ is the mass of the ghost subhalo, $\rho_{\rm h}$ is the density of the host at the location of the ghost subhalo, σ is the radial velocity dispersion of the DM particles (assumed to be Maxwellian) and $\ln\Lambda$ is the so-called Coulomb logarithm.

$M_{\rm s}$ is computed assuming that the subhalo is described by the same NFW profile it had when it was last detected truncated at the radius $R_{\rm t}$ introduced in Section 3.1. The host density is computed from the NFW profile $\rho_{\rm h}(R)$ of the host halo, where R is the distance of the ghost subhalo from the centre of the host. The radial velocity

dispersion σ is taken from the appendix A of Duarte & Mamon (2015).[5] The Coulomb logarithm is given by:

$$\ln\Lambda = \ln\left(1 + \frac{M_{\rm h}}{M_{\rm s}}\right).\tag{7}$$

3.3 Survival time

The problem of computing the survival time $t_{\rm surv}$ is that of determining after how many pericentric passages we can stop integrating equation (4) because we can consider that the satellite galaxy has merged with the central galaxy of the host halo. Our calculation is based on a modified version of the standard dynamical friction in Binney & Tremaine (2008), which we briefly rederive to clarify its assumption.

In the simplifying case of circular orbits, equation (4) implies that the specific angular momentum loss due to the dynamical-friction force is:

$$\frac{\rm d}{{\rm d}t}(R\, V_{\rm c}) = -R\, a_{\rm df}.\tag{8}$$

For a singular isothermal sphere, $V_{\rm c} = \sqrt{2}\sigma$ is independent of radius and $\rho_{\rm h} = V_{\rm c}^2/(4\pi G R^2)$. Hence, inserting equation (5) into equation (8) leads to:

$$\frac{{\rm d}R}{{\rm d}t} = -f(1)\ln\Lambda\,\frac{G\,M_{\rm s}}{R\,V_{\rm c}}.\tag{9}$$

The time the satellite takes to spiral in from $R = R_{\rm vir}$ to 0 is thus (Binney & Tremaine 2008):

$$t_{\rm df} = \int_0^{R_{\rm vir}} \frac{V_{\rm c}}{f(1)\ln\Lambda G M_{\rm s}} R\, {\rm d}R = \frac{A}{\ln\Lambda}\frac{M_{\rm h}}{M_{\rm s}}\frac{R_{\rm vir}}{V_{\rm c}},\tag{10}$$

with $A = 1/[2f(1)] \simeq 1.17$, since $V_{\rm c}^2 = GM_{\rm h}/R_{\rm vir}$.

On eccentric orbits, $\rho_{\rm h}$ varies on a time-scale $t_{\rm dyn} \ll t_{\rm df}$ invalidating equation (10) (see Mamon 1996; Chan, Mamon & Gerbal 1997; Cora, Muzzio & Vergne 1997).[6] Furthermore, in equation (10), we could take $M_{\rm s}$ out of the integral because we assumed it to be constant. Real subhaloes are stripped by the tidal field of the host. This reduces the dynamical-friction force and slows down the orbital decay (Mamon 1987; Yi et al. 2013).

Jiang et al. (2008) have investigated these effects with cosmological hydrodynamic simulations. They have found that equation (10) gives an accurate measure of the time-scale on which a satellite initially at $R = R_{\rm vir}$ merges with the central galaxy if the Coulomb logarithm is computed with equation (7) and if the coefficient $A = 1.17$ is replaced by:

$$A = 1.17\,(0.94\,\epsilon^{0.6} + 0.6),\tag{11}$$

where ϵ is the orbital circularity, that is, the ratio of the angular momentum to that of a circular orbit with the same total energy ($\epsilon = 1$ for circular orbit and $\epsilon = 0$ for radial orbit).

Moster, Naab & White (2013) modelled orphan galaxies/ghost subhaloes in a manner similar to ours. They used equation (10) with

---

[5] Duarte & Mamon (2015) solved the Jeans equation of local dynamical equilibrium for a velocity anisotropy $\beta = 1 - 2\sigma_r^2/\sigma_t^2$ with the radial dependence $\beta = (1/2)[1 - R/(R + R_0)]$, which Mamon & Łokas (2005) found to match well the velocity anisotropies measured in cosmological simulations (R_0 is the scale radius of the NFW profile).

[6] Even on circular orbits, the time-scale for orbital decay can be up to four smaller than predicted by the Chandrasekhar formula, because of resonances between the halo and subhalo (Prugniel & Combes 1992).
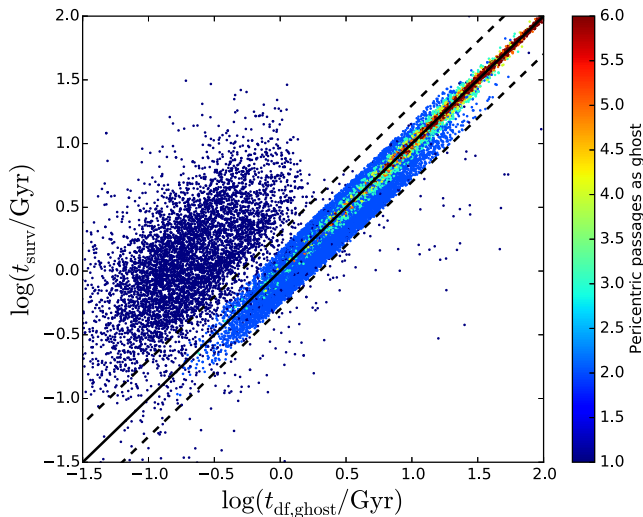
Figure 2. Comparison, for ghost subhaloes, between the expected survival time $t_{df, ghost}$ based on equations (10) and (11) and the actual survival time $t_{surv} = t_{merg} - t_{ghost}$ in our model, where mergers can occur only at a pericentric passage. For clarity, we show only 10 per cent of the ghosts of the simulation box. The *black solid line* represents $t_{surv} = t_{df, ghost}$. The *dashed lines* correspond to a scatter ± 0.3 dex and enclose 74 per cent of the points on the diagram. Points are colour-coded according to the number of pericentric passages between the subhalo's last detection and the merger time in our model. The fractions of ghost subhaloes that merge at the first and the second pericentric passage are 23 per cent and 40 per cent, respectively.

$A = 2.34$, this value being based on idealized simulations of orbital decay by Boylan-Kolchin, Ma & Quataert (2008). Given the mean circularity $\langle \epsilon \rangle \simeq 0.55$ found by Jiang et al., the mean value of A in equation (11) is 1.47. Hence, our dynamical-friction times are shorter than those used by Moster et al. by 40 per cent on average.

The dynamical-friction time t_{df} computed with equations (10) and (11) sets the initial value of the merging countdown time, which begins to tick for detected and ghost subhaloes at the time they first enter the virial radius. When the time t_{df} elapses, a subhalo can be at any point of its orbit (orbits shrink by dynamical friction but, for highly elongated orbits, the apocentre may still be far away from the centre of the host). Physically, however, the merger of a satellite galaxy with the central one is expected to occur at a pericentric passage. We therefore assume that a ghost subhalo merges with its host at the pericentric passage that is closest in time to when the merging countdown timer rings (marked by a red cross in the example of Fig. 1).

In formulae, let t_{entry} be the time at which the subhalo enters the virial radius for the first time (the large black circle in Fig. 1 shows R_{vir} at t_{entry}), let t_{ghost} be the time at which the subhalo ceases to be detected (the last red empty circle in Fig. 1) and let t_{merg} be the time of the pericentric passage at which the galaxy merger takes place (the point where the red curve ends). Then, $t_{df, ghost} \equiv t_{entry} + t_{df} - t_{ghost}$ is the remaining dynamical-friction countdown timer when the subhalo turns into a ghost subhalo. The survival time of the ghost subhalo from the time it becomes a ghost, $t_{surv} \equiv t_{merg} - t_{ghost}$, can be both larger or smaller than $t_{df, ghost}$ depending on whether the nearest pericentric passage occurs before or after the cosmic time $t_{entry} + t_{df}$. However, Fig. 2 shows that most ghost subhaloes (74 per cent) lie on a tight correlation $t_{surv} \simeq t_{df, ghost}$. Nearly all the outliers merge at their first pericentric passage. They are subhaloes that ceased being detected short after a pericentric passage and for which the merging countdown timer rang while

they were still detected. As we assume that mergers can occur only at pericentric passages, these subhaloes were obliged to make another orbit even though their merging countdown timer had come to zero.

4 THE ENTRY MASSES

This section explains our procedure to assign entry masses to galaxies that enter a group or cluster environment. The entry redshift z_{entry} is the redshift at which the subhalo associated with a satellite galaxy is detected as subhalo of its host for the first time, and differs for each galaxy. Since until z_{entry} all galaxies are central, we can assume that at z_{entry} galaxies still obey the stellar mass–halo mass relation for central galaxies, so that $m_{entry} = m_c(M_h, z_{entry})$, where m_c is the stellar mass of the central galaxy for a halo of mass M_h at z_{entry}.

In principle, we could derive $m_c(M_h, z_{entry})$ by solving equation (1), where n_* is the SMF of central galaxies at z_{entry} and n_h is the HMF (without subhaloes) at z_{entry}. In practice, while it is very easy for a theorist to measure n_h in an *N*-body simulation with or without subhaloes at any redshift, separating central and satellite galaxies in the observations is much more difficult: this requires large spectroscopic surveys and has only been done so far for local data (the Main Galaxy Sample of the SDSS, where nearly all galaxies lie at $z < 0.2$; Yang et al. 2009, 2012). The red dashed line in Fig. 3 shows the SMHM relation that we obtain when we apply the procedure described in this paragraph using Yang et al.'s (2012) measurement of the local SMF of central galaxies on the seventh data release of the SDSS.

The problem of this approach is that many of our satellite galaxies have z_{entry} outside the redshift range probed by Yang et al. (2012). Fig. 4 shows the cumulative distribution of z_{entry} for different bins of host–halo (group) mass. The median entry redshift is $z_{entry} \simeq 0.1$ for $M_h > 10^{13.5}\,M_\odot$ but $z_{entry} \simeq 0.3$ for $M_h < 10^{13.5}\,M_\odot$, while less than one satellite in a thousand has $z_{entry} > 2.5$. At $z > 0.2$, we only have the total SMF of galaxies, with no splitting between centrals
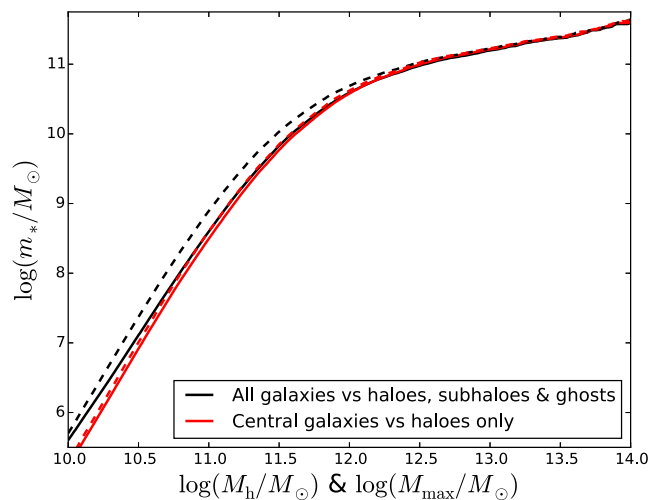
Figure 3. SMHM relation computed by AM (equation 1) using the local ($z < 0.2$) data of Yang et al. (2012). The AM is performed between the SMF of central galaxies and the HMF without subhaloes (*red*) or between the total SMF (central and satellite galaxies) to the total mass function of haloes and subhaloes, and ghosts (*black*). The HMF is computed from the virial mass M_h measured in the *N*-body simulation at $z \simeq 0.1$, using the procedure described in Section 2 (*dashed lines*) or from the maximum virial mass M_{max} that a halo/subhalo and its main progenitor ever had over its entire history at $z \gtrsim 0.1$ (*solid lines*).
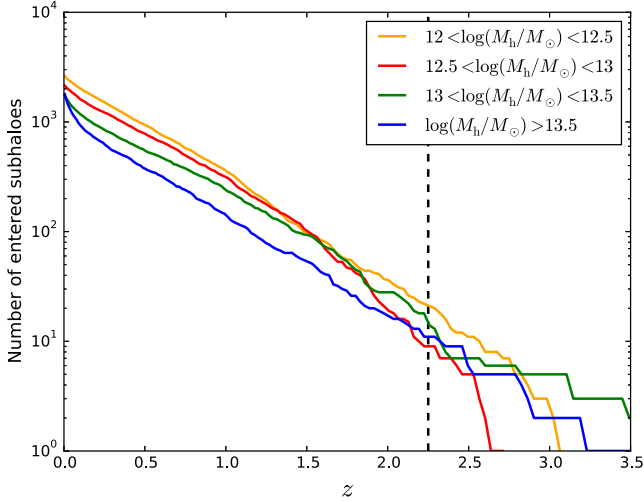
Figure 4. Cumulative distribution of halo entry redshifts for different bins of host-halo mass. The yellow, red, green and blue curves show the mean distribution $\mathrm{d}N/\mathrm{d}z$ for the entry redshift $z = z_{\mathrm{entry}}$ in the bin of host-halo mass $12 < \log(M_{\mathrm{h}}/\mathrm{M}_\odot) \le 12.5$, $12.5 < \log(M_{\mathrm{h}}/\mathrm{M}_\odot) \le 13$, $13 < \log(M_{\mathrm{h}}/\mathrm{M}_\odot) \le 13.5$ and $\log(M_{\mathrm{h}}/\mathrm{M}_\odot) > 13.5$, respectively. The vertical dashed line at $z = 2.5$ marks the upper boundary of the redshift range probed by Muzzin et al.'s (2013) data. At $z < 0.2$, entry masses are computed from the SMF of Yang et al. (2012), who also provided a central/satellite decomposition.

and satellites. To explore the consequences of approximating the SMF of central galaxies with the total SMF, we start by making this approximation in the local Universe, where we know the correct answer $m_{\mathrm{c}} = m_{\mathrm{c}}(M_{\mathrm{h}})$ given by the red dashed curve in Fig. 3. If n_* is the total SMF and n_{h} is the HMF including subhaloes and ghosts, then equation (1) gives the relation $m_* = m_*(M_{\mathrm{h}})$ shown by the black dashed curve in Fig. 3. The difference between the black and the red dashed curves is sufficiently large that approximating the latter with the former would compromise our analysis.

The black dashed curve lies above the red dashed curve because DM haloes are stripped more easily than the compact luminous galaxies at their centres. Subhaloes are stripped more heavily and have higher stellar-to-DM mass ratios m_*/M_{h} than haloes. To demonstrate that tidal stripping of subhaloes is the physical origin for the difference between the black and the red dashed curves, we have recomputed the curves using the maximum mass M_{max} that a halo ever had across its history rather than the virial mass M_{h} at $z \simeq 0.1$ as an estimator of M_{h}. In this definition, M_{max} cannot decrease. Thus, this procedure removes the effects of mass loss from haloes/subhaloes in our analysis. The relation $m_{\mathrm{c}} = m_{\mathrm{c}}(M_{\mathrm{max}})$ for central galaxies (the red solid line) is very similar to $m_{\mathrm{c}} = m_{\mathrm{c}}(M_{\mathrm{h}})$ (the red dashed line) because, for haloes, mass loss is usually negligible. However, the relation $m_* = m_*(M_{\mathrm{max}})$ for all galaxies (the black solid line) is significantly different from $m_* = m_*(M_{\mathrm{h}})$ (the black dashed line).

The main conclusion of Fig. 3 is that $m_*(M_{\mathrm{max}}) \simeq m_{\mathrm{c}}(M_{\mathrm{h}})$ for $M_{\mathrm{max}} = M_{\mathrm{h}}$ (the black solid line and the red dashed line are very close), at least for $M_{\mathrm{h}} > 10^{10.5}\,\mathrm{M}_\odot$ and $m_* > 10^7\,\mathrm{M}_\odot$. Thus, we are justified to replace our original assumption $m_{\mathrm{entry}} = m_{\mathrm{c}}(M_{\mathrm{h}}, z_{\mathrm{entry}})$ with $m_{\mathrm{entry}} = m_*(M_{\mathrm{max}}, z_{\mathrm{entry}})$, from which we can compute entry masses at redshifts much larger than $z = 0.2$.

The observational data that we use for the AM are the SMFs of Yang et al. (2012) at $z < 0.2$ and of Muzzin et al. (2013) at $0.2 < z < 2.5$. To avoid noise, we do the AM using four-parameter double-power-law fits to the observed SMFs rather than the data
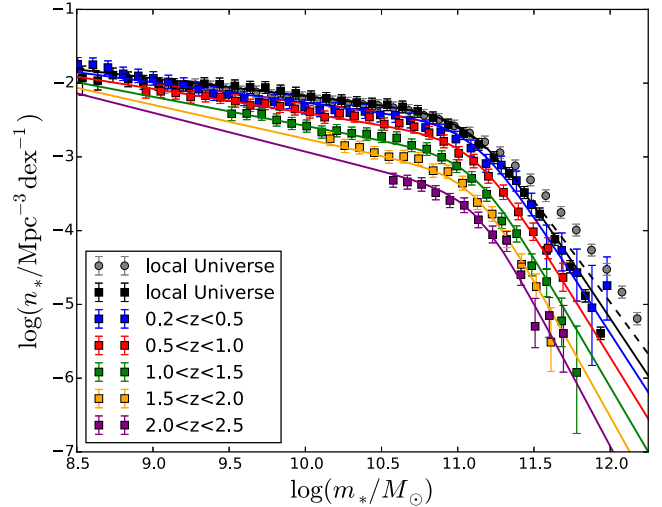

Figure 5. Observed SMFs used to compute the SMHM relation in Figs 3 and 6. The local data ($z < 0.2$, *black squares*) are from Yang et al. (2012). The data at $0.2 < z < 2.5$ (*blue, red, green, yellow* and *purple squares*) are from Muzzin et al. (2013). The SMFs in the six redshift bins were fitted with a double-power-law model (*curves*), the parameters of which were assumed to vary linearly with redshift. See Appendix C for more details about the fit. The *black, blue, red, green, yellow* and *purple curves* show our fits at $z = 0.1, 0.35, 0.75, 1.25, 1.75, 2.25$, respectively. The *grey circles* are data from Bernardi et al. (2013) ($z < 0.2$). They have not been used to fit the evolution of the SMF, but are shown for comparison. The *dashed black curve* is the fit we would have obtain by fitting Muzzin et al.'s (2013) data only and extrapolating them at $z \sim 0.1$. The quality of the fit to the black symbols and the small difference between the solid and dashed black curves proves the overall consistency of Muzzin et al.'s (2013) and Yang et al.'s (2012) data.

points themselves. Furthermore, to compute $n_*(m_*, z)$, we do not use the best-fitting parameters at redshift z. We determine the parameter value at redshift z by fitting a straight line to the evolution with redshift of the best-fitting parameters over six redshift bins covering the range $0 < z < 2.5$. The exact fitting formula and the values of the parameters used to fit the SMF are presented in Appendix C. Fig. 5 shows that this model for the galaxy SMF is in good agreement with the data points of both Muzzin et al. (2013) and Yang et al. (2012).

To assess the consistency of the two data sets, we have computed the local SMF by extrapolating the data of Muzzin et al. (2013) to $z \sim 0.1$, without including the data of Yang et al. (2012) in the fit. The result (the black dashed line in Fig. 5) is intermediate between the SMFs of Yang et al. (2012) and Bernardi et al. (2013), but much closer to the former than to the latter.

Fig. 6 summarizes the results of this section by showing the m_*–M_{max} relation from AM at three different redshifts ($z = 0, 1, 2$). The black curve ($z = 0$) is smooth up to $M_{\mathrm{max}} \simeq 10^{13.3}\,\mathrm{M}_\odot$, where the effects of low-number statistics in the N-body simulations begins to be important (there are not many clusters in a volume of $10^6\,\mathrm{Mpc}^3$). We also note that, for a fixed halo mass (e.g. $M_{\mathrm{max}} = 10^{11.5}\,\mathrm{M}_\odot$), m_* is higher at lower z, most likely because there has been more time to convert gas into stars.

In Fig. 6, we also compare our SMHM relation at $z = 0$ with lensing data (Leauthaud et al. 2010; Reyes et al. 2012, circles) and previous AM/HOD models (Behroozi, Wechsler & Conroy 2013; Leauthaud et al. 2012; Papastergis et al. 2012; Moster et al. 2013, curves). The agreement with lensing data is very good considering that lensing observations measure $M_{\mathrm{h}} \le M_{\mathrm{max}}$. The agreement with previous AM/HOD models is also quite good (particularly in the
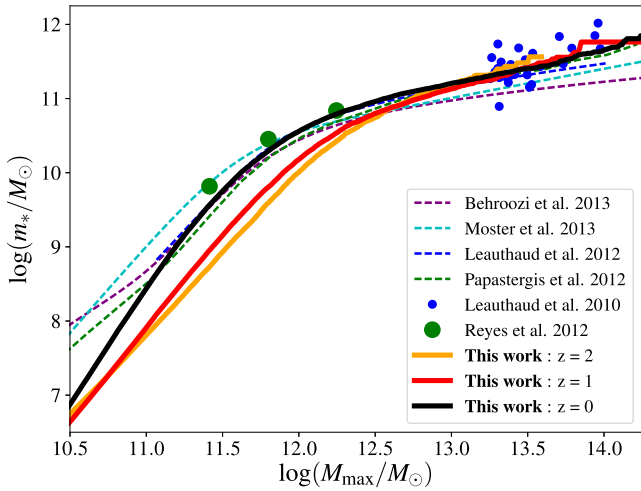
Figure 6. Stellar mass (m_*)–halo mass (M_{max}) relation used to compute m_{entry}. The relation is computed from AM at all time-steps in the merger tree. For clarity, we show it only at $z = 0$, 1 and 2 (thick solid black, red and orange curves, respectively). The thin dashed curves show the results of previous studies by Behroozi et al. (2013, purple), Leauthaud et al. (2012, blue), Papastergis et al. (2012, green) and Moster et al. (2013, cyan). The circles are lensing data. Each large green circle is the result of stacking ~10 000 spiral galaxies (Reyes et al. 2012), while the small blue circles are data points for individual galaxies (Leauthaud et al. 2010).

mass range $10^{11} \, M_\odot < M_{max} < 10^{12.5} \, M_\odot$), although models differ from one another at the level of 0.1–0.2 dex. The impact that this disagreement may have on our results is discussed in Section 7.1.5. To ease the comparison with future AM work, we provide in Appendix C a fit of our SMHM relation.

5 THE STELLAR MASSES OF SATELLITE GALAXIES TODAY

5.1 Evolution without tidal stripping

We have described the AM method that we used to place a galaxy at the centre of each halo in the volume of our N-body simulations. This procedure determines the stellar masses that satellite galaxies have when they enter a group or cluster environment. We now consider how the masses of these galaxies evolve after their haloes have become subhaloes. Here, we focus on the evolution without tidal stripping of stars, the effects of which will be discussed in detail in the next section.

In standard SAMs, a galaxy is composed of stars and cold gas, and is surrounded by a halo of hot gas. The halo of hot gas accretes mass from the intergalactic medium when the DM halo grows. The hot gas cools and accretes on to the galaxy. The cold gas within the galaxy forms stars. When the galaxy enters a larger system and becomes a satellite, the hot component associated with the galaxy can no longer grow and is depleted by ram-pressure stripping, tidal stripping or cooling on to the galaxy (see McCarthy et al. 2008 for a simple analytical model of how a subhalo is stripped of its hot gas). Only after the halo of hot gas has been stripped down to the size of the galaxy do ram-pressure and tidal stripping begin to remove the cold gas within the galaxy (Bekki 2009). Stars are the last to go because they are impervious to ram pressure and can be stripped only by tides.

Our model is focused on the stellar component and does not follow the presence of gas in either the hot or the cold component.

The stellar mass of a central galaxy is determined by an empirical relation that depends only on halo mass M_{max} and redshift, that is, cosmic time. Its growth is the sum of two terms, the growth of stellar mass with halo mass at constant cosmic time and the growth of stellar mass with time at constant halo mass:

$$\frac{\mathrm{d}}{\mathrm{d}t} m_*(M_{max}, t) = \left(\frac{\partial m_*}{\partial M_{max}}\right)_t \dot{M}_{max} + \left(\frac{\partial m_*}{\partial t}\right)_{M_{max}}. \qquad (12)$$

The first term is directly related to the accretion of baryons on to the halo. Thus, it is natural to interpret the second term as the depletion of an existing gas reservoir by star formation (keeping in mind that star formation is not the only process that may remove gas from galaxies). In central or isolated galaxies, the first term dominates. We say that these galaxies are in an 'accretion mode'. In satellite galaxies, the only contribution to the star formation rate comes from the second term. We say that these galaxies are in a 'starvation mode'. We note that our definition of starvation does not exclude accretion from a residual reservoir of hot gas. This definition may not coincide with that of other authors, who define starvation as a complete shutdown of gas accretion on to the galaxy.

The transition from an accretion mode to a starvation mode at z_{entry} is not an assumption of our model. It is a consequence of the fact that subhaloes do not gain mass, they lose it. For most subhaloes, $\dot{M}_{max} = 0$ at $z < z_{entry}$

In a more extreme scenario, the entire reservoir potentially available for star formation (hot and cold gas) is removed from satellite galaxies upon entry into the host halo. In this 'shutdown' scenario, $\dot{m}_* = 0$ for all satellite galaxies at $z < z_{entry}$. We call this scenario, the shutdown model because it corresponds to a complete shutdown of star formation in satellite galaxies. The shutdown and the starvation models set lower and upper limits, respectively, to the star formation that is possible in group and cluster environments.

The blue and the red curves in Fig. 7 illustrate the qualitative evolution of m_* in the starvation and the model, respectively, when we neglect the effects of tides. Tidal stripping transforms the blue curve into the green one and the red curve into the yellow one, but here we focus on models without tidal stripping because we postpone its discussion to Section 5.2.

In models without stripping, there is no mechanism that can remove stellar mass from galaxies, Hence at each time-step, the stellar mass of a central galaxy is updated to the maximum between the sum of the stellar masses of its progenitors[7] and the mass predicted by the m_*–M_{max} relation at the current z. However, requiring that the mass of a central galaxy be the largest between the sums of the stellar masses of its progenitors and the mass from AM method overestimates the masses of cD galaxies because the masses from AM already includes the effects of mergers. This assumption applies to both the shutdown and the starvation models.

We deal with this problem by introducing a maximum halo mass M_{lim}, above which we assume that dissipationless mergers are the only opportunity for galaxies to grow. M_{lim} is a free parameter to be determined by fitting the SMF of galaxies (Section 6.1). In haloes with $M_{max} > M_{lim}$, m_* is the sum of the stellar masses of the progenitors independently of what the AM relation prescribes. The condition that $m_*(M_{max})$ can never be lower than the value set by the AM relation is recovered in the limit $M_{lim} \to \infty$.

---

[7] By progenitors, we mean the galaxy main progenitor and all the satellites that have merged with it, since the last time-step.
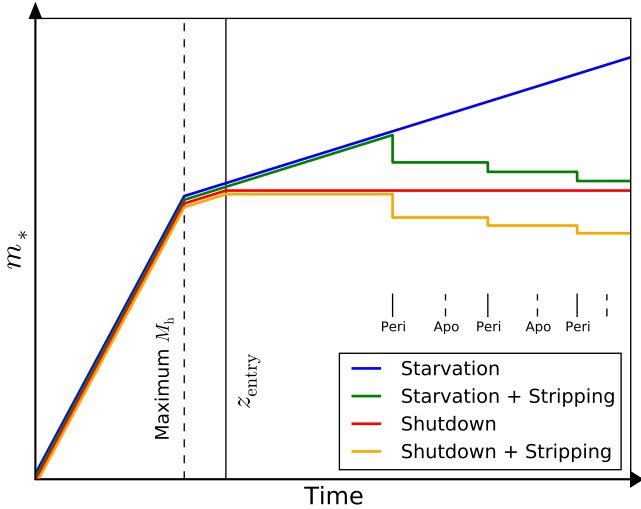
Figure 7. Illustration of our four scenarios for the evolution of stellar mass as a galaxy enters (at time z_{entry}, *black vertical bar*) and orbits (at times *peri* and *apo* for pericentres and apocentres, respectively) in a group or cluster (and thus transitions from a central in a small group before entry to a satellite in a larger group or cluster after entry therein). In all four models, the stellar mass first grows as expected from AM with the current halo maximum mass, once the (sub)halo mass reaches its maximum value (*black dashed vertical bar*), the stellar mass grows more slowly until z_{entry} is reached. In the shutdown model (*red*), the stellar mass after entry is maintained at the value at entry. In the starvation model (*blue*), the stellar mass increases following the AM prescription. Tidal stripping occurs at orbit pericentres (*orange and green* for shutdown and starvation, respectively). After the first stars have been stripped no more star formation is allowed. Satellite mergers are not considered in this illustration.

5.2 Tidal stripping of stellar mass

We now need to estimate the effects of tidal stripping on the stellar mass of the subhaloes. As a gravitational dynamical process, tidal stripping does not distinguish between a star and DM particle. Therefore, the tidal radius, r_t of the stellar distribution should match the tidal radius, R_t of the subhalo, which we computed in Section 3.1 for the DM, especially if we neglect the different dynamical structures of discs and haloes, which is beyond the scope of our analysis, particularly since we do not distinguish between satellites of different morphological types.

However, while a subhalo loses an important fraction of its DM mass before its first pericentric passage (Klimentowski et al. 2009; although we cannot exclude that this may be due to incomplete relaxation of the subhalo), the more concentrated stellar component is stripped almost entirely at pericentric passages (strong variations of the tidal acceleration along elongated orbits lead to a tidal shock at pericentre; Ostriker, Spitzer & Chevalier 1972 and fig. 3 of Klimentowski et al. 2009). Fig. 7 illustrates the qualitative effect of adding tidal stripping on elongated orbits to our shutdown (now orange) and starvation (now green) models. Since most satellite galaxies/subhaloes are on elongated orbits (Ghigna et al. 1998), an impulsive model for tidal stripping of stars is justified. In contrast, the circular-orbit approximation behind equation (3) leads to errors that are too large.

Moreover, equation (3) is based on the assumption that a particle is immediately stripped as soon as the tidal acceleration is larger

than the gravitational acceleration that keeps it bound to the satellite. In reality, it is also necessary that the impulse

$$\Delta \boldsymbol{v} = \int \boldsymbol{a}_t \, dt \tag{13}$$

imparted by the tidal acceleration $\boldsymbol{a}_t$ to the particle be sufficient for the unbinding condition

$$\frac{1}{2}(\boldsymbol{v}_s + \Delta \boldsymbol{v})^2 + \Phi_s \geq 0 \tag{14}$$

to be satisfied, where $\boldsymbol{v}_s$ is the velocity of the particle in the satellite before the tidal perturbation and Φ_s is the gravitational potential of the satellite system.

The pericentric passage is where the tidal acceleration a_t is strongest, but also where the passage is fastest. The impulsive approximation consists of assuming that the integral in equation (13) is dominated by the contribution around the pericentre. This leads to

$$\Delta v \sim a_t \, \Delta t_p \sim |\alpha| \frac{GM_h(R_p)}{R_p^3} \left(\frac{R_p}{V_p} \right) r \,, \tag{15}$$

where $\Delta t_p \sim R_p/V_p$ is the duration of the pericentric passage (R_p and V_p are the pericentric radius and speed of the satellite in the host's reference frame), $M_h(R_p)$ is the host-halo mass within R_p, r is the distance of the particle from the centre of mass of the satellite, α is the exponent of the mean-density profile and a_t has been evaluated with equation (A11) from Appendix A. (Similar equations were derived by Spitzer 1958 for a point mass perturber, and generalized to extended perturbers by Gonzalez-Casado, Mamon & Salvador-Sole 1994 and Mamon 2000.)

Let us assume that $\langle \boldsymbol{v}_s \cdot \boldsymbol{a}_t \rangle = 0$, either because $\boldsymbol{v}_s$ has a random orientation or because the particle is assumed to lie on a circular orbit (as stars in the discs of spiral and S0 galaxies), in which case the only component of $\boldsymbol{a}_t$ parallel to $\boldsymbol{v}_s$ is the azimuthal one, which vanishes when averaged over the orbit (Appendix A). Then, as the term $\boldsymbol{v}_s \cdot \Delta \boldsymbol{v}$ in equation (14) vanishes, substituting equation (15) into equation (14) gives:

$$\frac{1}{2} \left[|\alpha| \frac{GM_h(R_p)}{R_p^2} \frac{r_t}{V_p} \right]^2 = -\Phi_s(r_t) - \frac{1}{2} v_s^2. \tag{16}$$

With the substitution $V_c^2(R_p) = GM_h(R_p)/R_p$, equation (16) becomes

$$\frac{\alpha^2}{2} \frac{GM_h(R_p)}{R_p^3} \frac{V_c^2(R_p)}{V_p^2} r_t^2 = -2 \, \Phi_s - v_s^2 \,. \tag{17}$$

If we make the further assumption that the test particle is on a circular orbit, so that $v_s^2(r_t) = GM_s(r_t)/r_t$ (as expected for a star in the disc a spiral galaxy), then equation (17) can be rewritten in its final form:

$$\left[-2 \frac{\Phi_s(r_t)}{v_s^2(r_t)} - 1 \right] \frac{M_s(r_t)}{r_t^3} = \frac{\alpha^2}{2} \left[\frac{V_c(R_p)}{V_p} \right]^2 \frac{M_h(R_p)}{R_p^3}, \tag{18}$$

where R_p and V_p are computed from the conservation of energy:

$$\frac{1}{2} V_p^2 + \Phi_h(R_p) = e \tag{19}$$

and the conservation of angular momentum:

$$R_p V_p = j, \tag{20}$$

where e is the specific mechanical energy of the satellite, j is the specific angular momentum and Φ_h is the gravitational potential of the host system. Although e and j are not really conserved because

the satellite is subject to the dynamical-friction drag force (equation 4), their variations between snapshots are small and the values used to solve equations (19) and (20) are those measured at the snapshot just before the pericentric passage, where $\dot{R}$ changes sign.

Equation (18) is identical to equation (3), except that the term $|\alpha|$ is now replaced by:

$$\epsilon_{ts} = \frac{\alpha^2/2}{-2\Phi_s(r)/v_s(r)^2 - 1} \left[\frac{V_c(R_p)}{V_p} \right]^2.$$

(21)

To compute how ϵ_{ts} depends on the distance r from the centre of the satellite, we assume that all galaxies have an exponential scalelength (Mo, Mao & White 1998):

$$r_d = \lambda R_{vir}/2,$$

(22)

the value of which can neverdecrease. Here, λ is the spin parameter measured in the N-body simulation. At the denominator of equation (22), there is 2 rather than $\sqrt{2}$ because λ is defined as in Bullock et al. (2001). The exponential profile is truncated at the radius r_t computed with equation (18). All stars outside r_t are removed from the galaxy and re-assigned to the ICL without modifying the satellite's profile inside r_t. The assumption that all satellite galaxies are discs is admittedly extreme but we do not expect that this assumption significantly affects any of our results.[8]

Although the ICL is, by definition, intracluster, i.e. it is not associated with individual galaxies, we can imagine that, in the beginning, stars stripped from galaxies will form tidal tails, which can still be associated with the galaxies from which they originated (e.g. the Magellanic Stream and the Magellanic Clouds). Only later will phase mixing transform these tails into the extended envelopes of central galaxies. We thus begin by storing the stars stripped from individual galaxies in an ICL component that is still associated with the satellites from which it came. When the satellites merge, we transfer the stellar mass in this component to the ICL associated with the outer envelopes of the central galaxy.

Fig. 8 shows the radial dependence of the factor $\alpha^2/2/(-2\Phi_s/v_s^2 - 1)$, which sets the maximum efficiency of tidal stripping, for pure DM configuration (black solid curve) and when a disc is embedded in the subhalo (red curve). The figure shows that stripping is less efficient for stars in the central parts of a satellite galaxy. It also shows that $\epsilon_{ts} \ll 3$ everywhere.

Besides the coefficient in Fig. 8, the multiplicative factor $[V_c(R_p)/V_p]^2 < 1$ is the only significant difference between the results of the impulsive approximation (equation 18) and the instantaneous approximation (equation 3).

The dependence of the tidal radius on R_p and V_p was first proposed by Gonzalez-Casado et al. (1994) and confirmed with N-body simulations by Ghigna et al. (1998). Mamon (2000) considered a similar model, in which he assumed $\Phi_s \sim -GM_s(r_t)/r_t$, and found a simpler version of equation (21) without the term in Fig. 8 in front of the square bracket. He obtained the following formula, where V_c is the circular velocity of the host halo and v_c the circular velocity of the subhalo.

$$r_t \approx \frac{V_p}{V_c(R_p)} \frac{v_c(r_t)}{V_c(R_p)} R_p$$

(23)

Fig. 9 compares the tidal radii for our impulsive model (equation 21; solid curves), the model of Mamon (dashed curves) and the

[8] The satellite population is dominated by disc morphologies: spirals, S0s (the spiral arms are no longer visible because all gas has been consumed or removed) and dSphs (discs that puffed up because of either stellar feedback or tidal interactions).
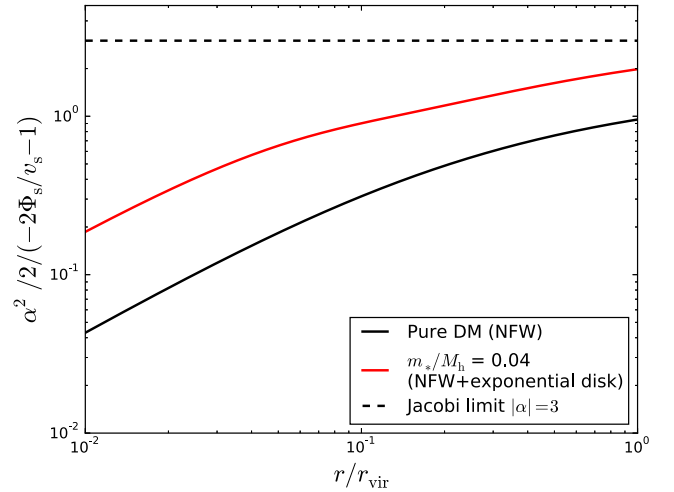

Figure 8. Maximum efficiency of tidal stripping in the impulsive approximation as a function of the distance from the centre of the satellite (r_{vir} is the virial radius of the satellite and $\alpha = -3$; the real efficiency is the maximum efficiency times $[V_c(R_p)/V_p]^2$). The *black curve* shows the radial dependence of the maximum efficiency for an NFW potential with $c = 8$. The *red curve* shows the effect of embedding in the subhalo a disc of mass $0.04 M_h$. These values correspond to the maximum m_{stars}/M_h ratio allowed by AM and to $\lambda = 0.05$, respectively. The horizontal *dashed black line* correspond to a circular orbit in the Jacobi limit ($|\alpha| = -3$).

instantaneous-tide circular-orbit approximation (filled symbols). By definition, the circular tidal theory is only valid for circular orbits, whereas the impulsive theory is not valid for circular orbits. Hence, we expect the tidal radii computed from the circular theory to be more accurate for orbits with $V_p = V_c(R_p)$ and the tidal radii computed from the impulsive theory to be more accurate for orbits with $V_p \gg V_c(R_p)$. However, most satellites have elongated orbits. The open symbols in Fig. 9 show the mean elongations, measured by $V_p \gg V_c(R_p)$, for orbits with different pericentric radii R_p/R_{vir}, according to Ghigna et al. (1998). Smaller pericentric radii correspond to higher orbital elongations. Comparing the ordinates of the open symbols to those of the filled ones shows that the circular theory underestimates the average tidal radius, particularly for satellites with small pericentric radii, and thus overestimates tidal stripping. Fig. 9 also shows that the simple theory of Mamon (2000) provides a good estimate of the tidal radius for satellites with $R_p \sim 0.1 R_{vir}$, although it overestimates the effects of tides for satellites on highly elongated orbits.

We therefore compute the tidal radius r_t for the stellar component with equation (18), which is more accurate. The implications for our results of using equation (18) rather than equation (3) will be discussed in Section 7.1.3.

The best way to test the validity of our model for tidal stripping is to compare it with idealized (i.e. non-cosmological) numerical simulations, which follow the dynamics of stars more accurately than our analytic calculations while retaining full control of the gravitational potential and the orbital configuration. Kazantzidis, Łokas & Mayer (2013) used idealized simulations to study how tidal stirring can transform a dwarf irregular (a discy dwarf) into a dwarf spheroidal. They assumed a host with the size of the Milky Way and explored three orbital configurations: $(R_a, R_p) = (125, 25)$ kpc, $(R_a, R_p) = (125, 50)$ kpc and $(R_a, R_p) = (250, 50)$ kpc, where R_a and R_p are the apocentric and the pericentric radii, respectively. In the simulations with $R_p = 50$ kpc, tidal stripping was negligible. For similar orbital configurations, stripping is negligible in our model,
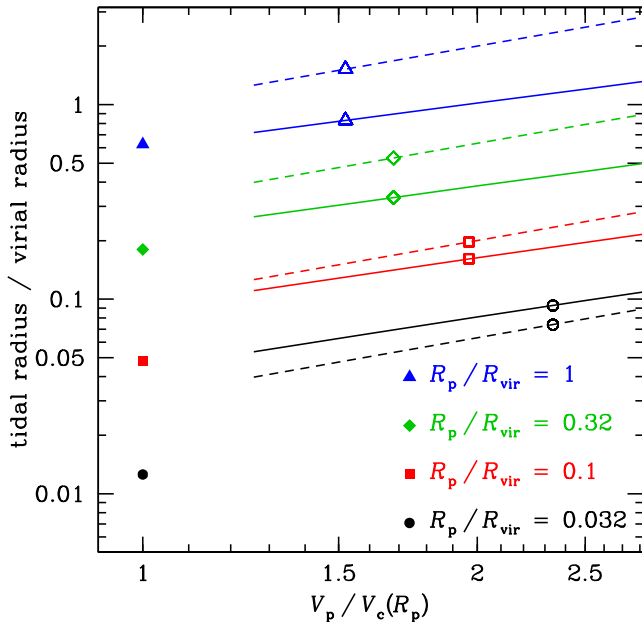
Figure 9. Tidal radii of subhaloes for different orbital elongations (measured by the ratio of pericentre speed to circular velocity at pericentre) and different pericentric radii (colours correspond to different values of R_p/R_{vir}). This figure compare the estimates from circular tidal theory (equation 3 with $\alpha = -3$, *filled symbols*) and from impulsive tidal theory using our full model (equation 18, *solid lines*) or the approximate formulation of Mamon (2000) (equation 23, *dashed lines*). The lines stop at $V_p/V_c(R_p) = 1.35$ because it makes no sense apply the impulsive theory to nearly circular orbits. The *open symbols* on the solid and dashed lines show the mean $V_p/V_c(R_p)$ for a given R_p/R_{vir}, computed assuming an average apocentre-to-pericentre ratio of five (Ghigna et al. 1998). As the only purpose of this figure is to compare different approximations, we have assumed that both the host and the subhalo are described by $c = 8$ NFW models and we have neglected the influence of the baryons on tidal radii.

too. In the simulation with $R_p = 25$ kpc, Kazantzidis et al. found the same qualitative behaviour that is shown by the models with stripping in Fig. 7. Quantitatively, the stellar mass of the satellite decreased by 5–10 per cent at each pericentric passage if the stripping potential corresponds to that of an NFW profile, as it does in our model. This figure is broadly consistent with our results, in which a satellite galaxy loses ∼20–25 per cent of its stellar mass over two pericentric passages on average (Section 6).

A more careful examination shows that the agreement is not so straightforward. If we apply our model of tidal stripping to the simulations of Kazantzidis et al. (2013), assuming the same stellar mass and radius as Kazantzidis et al., our model predicts the stellar mass stripped at the first pericentric passage is ∼1 per cent rather than the 5 per cent value found by Kazantzidis et al. in the simulation with $(R_a, R_p) = (125, 25)$ kpc. The stellar mass stripped at each pericentric passage is highly sensitive to both R_p and the radius of the satellite galaxy. The reason why tidal stripping is not negligible in our model despite being much weaker than suggested by Kazantzidis et al. is that the stellar component is less concentrated in our galaxies than in the dwarf of Kazantzidis et al. However, the comparison should keep in mind that Kazantzidis et al. defined m_* as the stellar mass within 0.7 kpc, corresponding to 1.7 exponential radii of the disc of the satellite galaxy. Observationally, disc galaxies in the central regions ($r < 0.1r_{200}$) of clusters tend to have surface-brightness profiles with residuals above an exponential fit at large radii (they have a type III 'antitruncated' profile; Pranger

et al. 2017). Pranger et al. interpreted this observation as a tidal effect. Instead of truncating discs, tides cause them to be more extended by pulling their outer regions. A lot of the stellar mass that Kazantzidis et al. considered as lost because it moved out of the central 0.7 kpc may still be in the disc at slightly larger radii. Our condition for stripping is much stronger because it requires gravitational unbinding (equation 14). Thus, it is not surprising that we find less stripping in our model. The interesting question is what stellar mass loss Kazantzidis et al. would have found if they had defined m_* as the total mass of the stars that are gravitationally bound to the satellite. We have no access to their simulations to answer this question.

6 RESULTS

6.1 Total SMF

In Section 5, we have described two models, the shutdown model and the starvation model, each in a version without and a version with tidal stripping. We now compare these four models with the local galaxy SMFs by Yang et al. (2012) and Bernardi et al. (2013; Fig. 10).

At $m_* \lesssim 3 \times 10^{11}$ M$_\odot$, the four models are indistinguishable from one another and they are all in excellent agreement with the SMF by Yang et al. (2012). However, at $m_* \gtrsim 3 \times 10^{11}$ M$_\odot$, all models tend to be above the SMF of Yang et al. The tendency is stronger for the starvation model without stripping than for the other three models. This finding may seem surprising because the starvation model without stripping applies to all haloes and subhaloes, an AM procedure that should reproduce the SMF of Yang et al., by construction. The discrepancy arises because, if the stellar mass returned by this procedure is smaller than the sums of the stellar masses of the progenitors of a galaxy, it is this sum and not the value returned by the AM procedure that is used to assign a stellar mass to this galaxy. If we assign to all galaxies a mass m_* such that the SMF of Yang et al. is reproduced, by construction, and we increase the masses of some of these galaxies (typically, the most massive ones, which have greatest number of progenitors), logically our SMF will contain more massive galaxies than the one of Yang et al. In other words, if all our haloes had a single progenitor, the blue curve in Fig. 10 would fit the square symbols by construction. The discrepancy between the blue curve and square symbols is linked to the merging histories of galaxies. The stellar mass above which the blue curve begins to differ from the SMF of Yang et al. (2012) is the one above which dry (dissipationless) mergers become the dominant growth mechanism and star formation is negligible (Cattaneo et al. 2011; Bernardi et al. 2011a,b).

By enforcing the AM relation, $m_* = m_*(M_{max})$ even when m_* is larger than the sum of the stellar masses of the progenitors, we effectively allow star formation in massive galaxies, which we know to be red from observations (Kauffmann et al. 2003; Baldry et al. 2004). We can deal with this problem by introducing a halo mass limit M_{lim}, above which m_* is simply the sum of the stellar masses of the progenitors, independently of the AM relation (Section 4). This is equivalent to assuming that there is a limit mass, above which dry mergers are the only growth mechanism.

Observations (Kauffmann et al. 2003; Baldry et al. 2004), physical models (Dekel & Birnboim 2006) and SAMs (Bower et al. 2006; Cattaneo et al. 2006; Croton et al. 2006) suggest a limit mass of order 10^{12} M$_\odot$ (Cattaneo et al. 2006 fitted the colour–magnitude distribution in the SDSS with $M_{lim} \sim 10^{12.4}$ M$_\odot$). We were therefore surprised to discover that the starvation plus stripping model
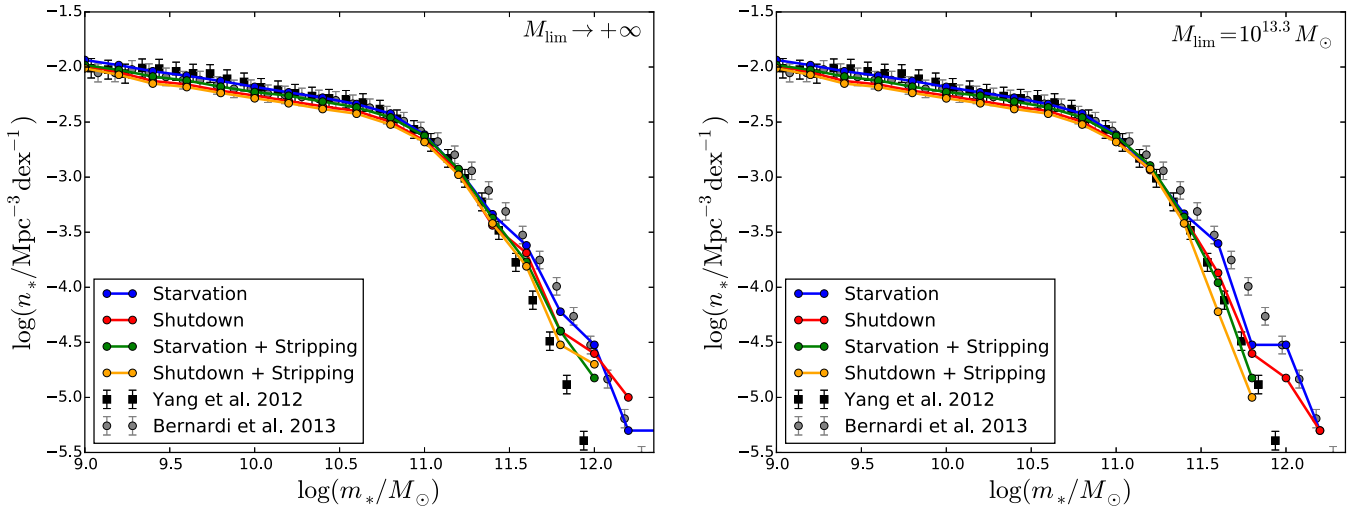
Figure 10. SMFs at $z = 0$ predicted by our four models (*curves*, colour-coded as in the legends and in Fig. 7) compared with the observations of Yang et al. (2012; *black squares* with error bars) and Bernardi et al. (2013; *grey circles* with error bars). Left: the stellar mass is always the maximum between the stellar mass from AM and the sum of the stellar masses of the progenitors. Right: when the halo mass is $M_{\rm max} > M_{\rm lim} = 10^{13.3}$ M$_\odot$, the stellar mass is the sum of the masses of the progenitors even when this mass is lower than the value obtained from AM.

fits the Yang et al. (2012) data for $M_{\rm lim} = 10^{13.3}$ M$_\odot$ (Fig. 10b, green curve). Our explanation for this finding is that AM is not a physical model for the baryonic mass that is able to condense to the centre in a halo of given mass: m_* is the net result of star formation, dry mergers and stripping. Tidal stripping is responsible for the extended envelopes of giant ellipticals, the masses of which are underestimated by studies based on magnitudes from the SDSS pipeline (Bernardi et al. 2017). Had we calibrated the m_*–$M_{\rm max}$ relation on the SMF of Bernardi et al. (2013), which includes the light from the outer regions and therefore the debris of tidally disrupted satellites, the green curve in Fig. 10(b) would have shifted to higher masses by an amount comparable to the difference between the SMFs of Yang et al. (2012) and Bernardi et al. (2013). To bring the green curve back on the data points of Yang et al. (2012) would have then required $M_{\rm lim} \lesssim 10^{12.6}$–$10^{12.7}$ M$_\odot$ (i.e. the mass limit that shifts the blue curve on the black squares in our calibration), in better agreement with previous studies.

6.2 Conditional SMF

The conditional SMF $N(m_*|M_{\rm h})$ is defined so that $N(m_*|M_{\rm h})\,{\rm d}m_*$ is the average number of galaxies with mass between m_* and $m_* + {\rm d}m_*$ in a host halo of mass $M_{\rm h}$ (we have omitted the dependence on z because, in this section, we are only interested in the local Universe). It can be split into the contributions of central and satellite galaxies, in which case the former integrates to unity (there is only one central galaxy per halo).

In this section, we compare the four models in Fig. 10(b) to the conditional SMF measured by Yang et al. (2012). A meaningful comparison requires: (i) that we apply their same definition of host-halo mass and (ii) that we apply their same criterion to decide which galaxies belong to a group or cluster.

We start with point (i). Yang et al. (2012) did not measure $M_{\rm h}$ dynamically. They inferred the group mass M_h from the group total luminosity L by using the AM relation:

$$\int_L^\infty n_{\rm gr}(L')\,{\rm d}L' = \int_{M_{\rm h}}^\infty \tilde{n}_h(M_{\rm h}')\,{\rm d}M_{\rm h}', \qquad (24)$$
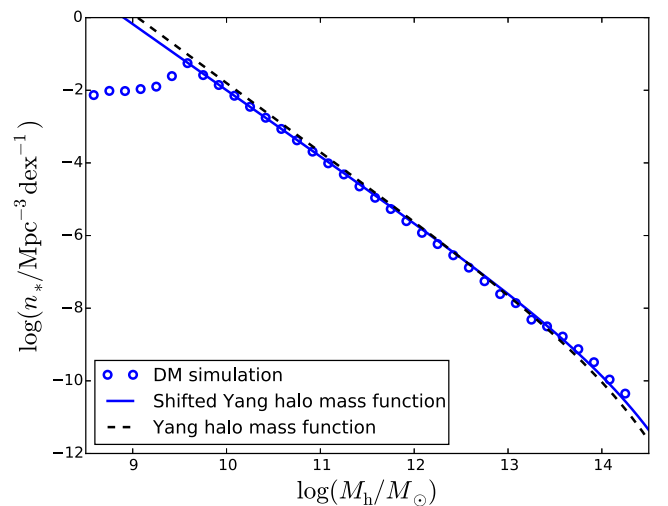
Figure 11. HMF assumed by Yang et al. (2012; *black dashed curve*) compared to the one that we extract from our *N*-body simulation (*blue circles*). The Yang et al. HMF (HMF) is mapped into our extracted HMF with a linear transformation $\log M_{\rm h} \mapsto a \log M_{\rm h} + b$, where a and b are fit to our HMF, and the best-fitting HMF is shown as the *blue curve*. The blue circles show clearly the resolution of our *N*-body simulations, which contains 1024^3 particles in a comoving volume of $(100\,{\rm Mpc})^3$.

where $n_{\rm gr}$ and $\tilde{n}_{\rm h}$ are Yang et al.'s (2012) group luminosity function and HMF, respectively. The problem is that the HMF $\tilde{n}_{\rm h}$ that they computed with Sheth & Tormen's (2002) formula (the black dashes in Fig. 11) is different from the HMF $n_{\rm h}$ that we measure in our *N*-body simulation (the blue open circles), also because the cosmology is not identical. To overcome this problem, we have fitted a linear transformation that maps $\tilde{n}_{\rm h}$ into $n_{\rm h}$ (i.e. the black dashes into the blue solid line). We have applied this transformation to the intervals of $M_{\rm h}$ within which Yang et al. (2012) determined the conditional SMF and we have used the transformed intervals to select host haloes of corresponding mass in our *N*-body simulation.

For point (ii), we have re-analysed the density profiles of Yang et al.'s (2012) groups and verified that they are truncated at R_{180}, the

radius within which the mean density equals 180 times the mean density of the Universe. R_{180}/R_{vir} depends on concentration. We have computed R_{180} for all haloes in the N-body simulations and used this radius to decide which satellite galaxies should be assigned to a host when computing the conditional SMF.

Fig. 12 compares the conditional SMF in our four models with Yang et al.'s (2012) data after taking points (i) and (ii) into account. A number of conclusions can be drawn from this comparison.

First, the difference between the blue curve and the green one (or the red curve and the orange one) is usually smaller than the difference between the blue curve and the red one. In other words, the effects of stripping are smaller than the uncertainty from our ignorance of the stellar mass Δm_* formed after t_{entry}.

Second, all our models predict an excess of massive satellites in low-mass groups ($M_h < 10^{13.44}$ M$_\odot$), though, at $m_* < 10^{10.5}$–10^{11} M$_\odot$, data points for satellites tend to lie in the range allowed by our models (between the blue and the orange curves).

Third, the starvation model with tidal stripping (green curves) is the one that, despite this problem, is overall in best agreement with the conditional SMF of Yang et al. (2012) (our N-body simulation contains very few clusters; therefore, the last two panels in Fig. 12 are affected by poor statistics). Fig. 12 was plotted for $M_{lim} = 10^{13.3}$ M$_\odot$, but these conclusions are based on the conditional SMF of *satellite* galaxies, the masses of which are insensitive to the value of M_{lim}.

Fig. 13 compares predictions and observations for the conditional SMFs displayed in Fig. 12 in a more quantitative manner. For each bin of group mass M_h, we compute the mean stellar mass of the central galaxy and the mean total mass of all satellite galaxies, and compare these masses to observations by taking their logarithmic differences. A model in perfect agreement with the observations would coincide with the black horizontal line $\log m_*^{model} - \log m_*^{obs} = 0$ everywhere. M_{lim} has been tuned so that the models with tidal stripping match well the observations at all stellar masses and group masses, but results for central galaxies (*dashed curves*) at $M_h < 10^{13}$–$10^{13.5}$ M$_\odot$ and satellite galaxies (*solid curves*) at all masses are insensitive to the value of M_{lim}. The solid curves should be interpreted with caution because they are the result of a sum over all m_*. For instance, the shutdown+stripping model (orange curve) is above the data at $M_h \sim 10^{12.5}$ M$_\odot$ because of a few massive satellites, while the conditional SMF for the corresponding model is below most data points in the mass bin $12.29 < \log (M_h/M_\odot) < 12.57$ (Fig. 12). However, there are two considerations that we can make from the solid curves in Fig. 13:

(i) Tidal stripping reduces the total stellar mass of satellite galaxies by typically 0.1 dex (0.2 dex at most, green versus blue and orange versus red curves).

(ii) The starvation model with tidal stripping (green curve) and the shutdown model without tidal stripping (red curve) provide a comparably good fit to the total stellar mass of satellites in a group.

Therefore, if we were to draw a conclusion based on the total stellar mass of satellites alone, we should concede that there is a degeneracy between the gas mass that accretes on to galaxies and the stellar mass that is stripped from them, and that a model with stellar mass loss intermediate between the predictions of the shutdown and the starvation models (between the red and the blue curves) could fit the observed conditional SMF without the need for any tidal stripping. Nevertheless, tidal stripping is expected to occur on physical grounds. Furthermore, there is observational evidence outside this work that the shutdown of star formation in satellite galaxies is not instantaneous (see the discussion in Section 7.1.2, and references

therein). Hence, it is re-assuring that the most astrophysically plausible model is the one that returns a comparatively best fit to the data in Figs 10 and 11.

6.3 ICL

As an independent test of our model, we have compared our predictions for the contribution of cD galaxies (inclusive of the ICL) to the total stellar masses of clusters with the observations of Gonzalez et al. (2013). Fig. 14 shows that, although there are very few clusters in our N-body simulations, our starvation plus tidal stripping model (green circles) matches the observed trend (crosses) for the ratio of BCG+ICL stellar mass over total stellar mass.

More interesting (but more difficult to compare with observations) is the contribution of the ICL to the total stellar mass within a cluster. Fig. 15 shows this contribution when we consider not only the extended envelops of cD galaxies but all stars stripped from galaxies over the entire cluster out to R_{vir}. It shows that stellar mass fraction in the ICL increases with halo mass.

In Fig. 14, we had shown the BCG+ICL mass fraction within R_{500} (the radius of a sphere within which the mean density equals 500 times the critical density of the Universe) for consistency with Gonzalez et al. (2013). In our cosmology, the virial radius corresponds to $\Delta_c = 102$ at $z = 0$, but we find that the ratio of the ICL mass m_{ICL} to the total stellar mass m_* is very similar within R_{vir} and R_{500} for haloes up to $\sim 10^{13.5}$ M$_\odot$. In clusters, m_{ICL}/m_* is ~ 20 per cent smaller within R_{500} than it is within R_{vir}. The implication is that the ICL is more concentrated than the total light of the cluster.[9] Fig. 15 shows that, in clusters, the ICL, defined as the total stellar mass stripped from galaxies, whether it still forms a tidal stream around the galaxies themselves or whether it has merged into the extended envelope of a cD galaxy, may amount to nearly half of the total stellar mass with R_{vir}.

In an article that appeared when ours was about to be submitted, Bernardi et al. (2017) argued against the interpretation that difference between the SMFs of Baldry et al. (2012) and Bernardi et al. (2013) is due to the ICL. Their claim is that the difference is entirely due to the different way the photometry is done. The magnitude provided by the SDSS are based on fitting an exponential and a de Vaucouleurs surface-brightness profile separately and retaining the value for the profile that fits better. The PYTHON image-morphology software PYMORPH fits the surface-brightness profiles of galaxies much more accurately because it allows for the presence of both an exponential and a Sérsic component. Bernardi et al. (2017) correctly argued that the difference is more than semantic because there is no doubt that a model with five free parameters can fit the surface-brightness profiles of galaxies more accurately than a model with two and thus return more accurate photometry.

As a Sérsic-exponential profile provides an excellent fit to the surface-brightness profiles of luminous red galaxies out to eight effective radii (about 100 kpc), Bernardi et al. (2017) concluded that the difference between PYMORPH and SDSS magnitudes cannot be due to the ICL. This conclusion is based on the fact that they define the ICL as any residual luminosity above the Sérsic-exponential fit. This definition is entirely reasonable from an observers

---

[9] Stripped stars are stored first in an ICL component associated with it parent satellite, when the satellites merge, we transfer the stellar mass in this component to the ICL associated with the central galaxy. Therefore, we define m_{ICL} within R_{vir} (respectively R_{500}) as the sum of the mass of the ICL component from galaxies within R_{vir}.
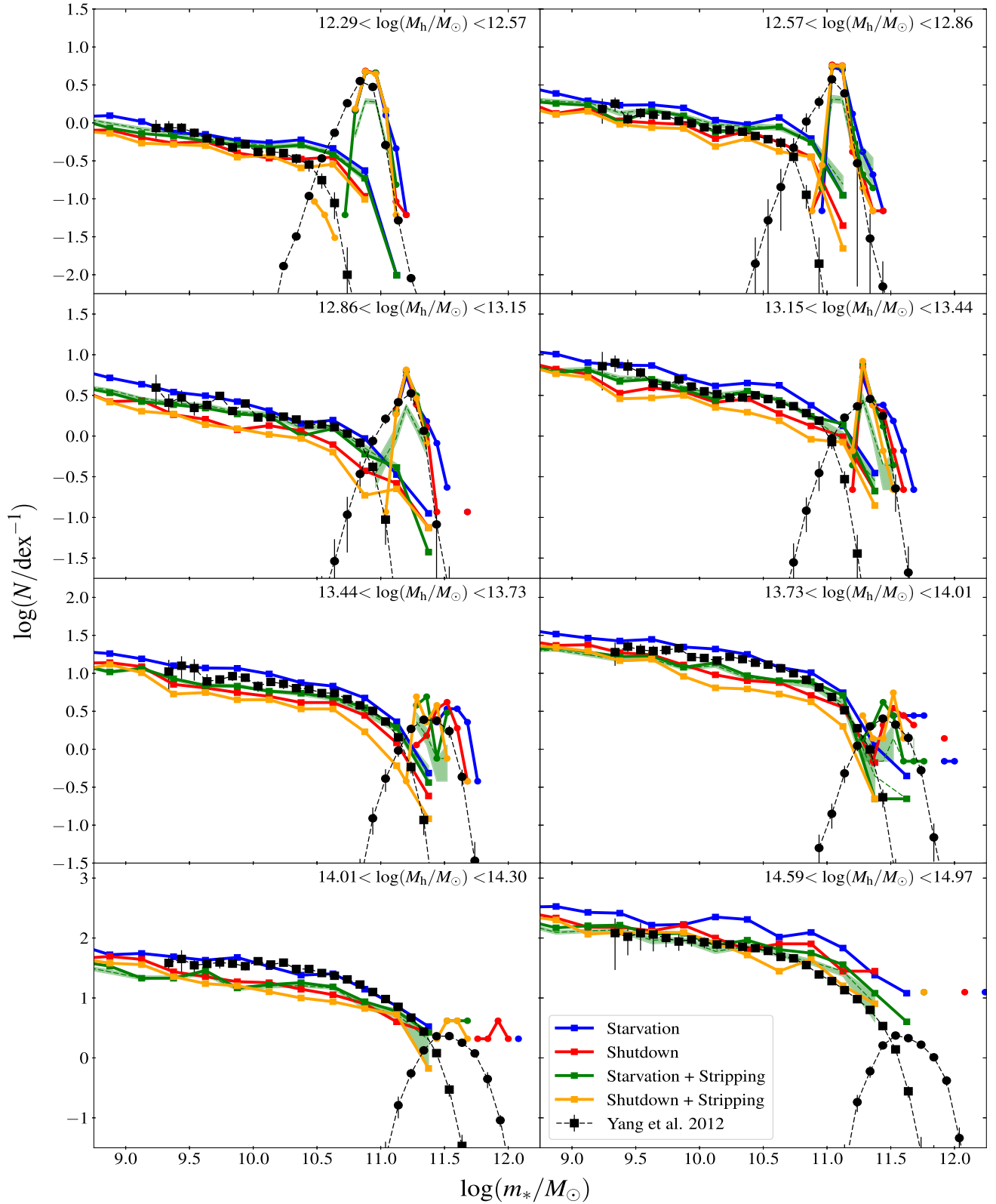
Figure 12. Conditional SMF for our different models (*thick solid curves*, see legends and Fig. 7) compared to the observations of Yang et al. (2012; *black points* with error bars). The panels correspond to bins of group mass. *Squares* and *circles* show Yang et al.'s decomposition of the data in central and satellite galaxies. The same decomposition has been applied to the models. In models with stripping (*green* and *orange curves*), the tidal radius r_t has been computed with equation (18). The thin green dashed curves show how the green thick curves vary when we introduce a scatter of 0.2 dex in SMHM relation. They are medians over a hundred realizations. The upper and lower envelopes of the green shaded areas around them correspond to upper and lower quartiles, respectively. In all panels, $M_{\rm lim} = 10^{13.3}$ M$_\odot$ (see Section 6.1).
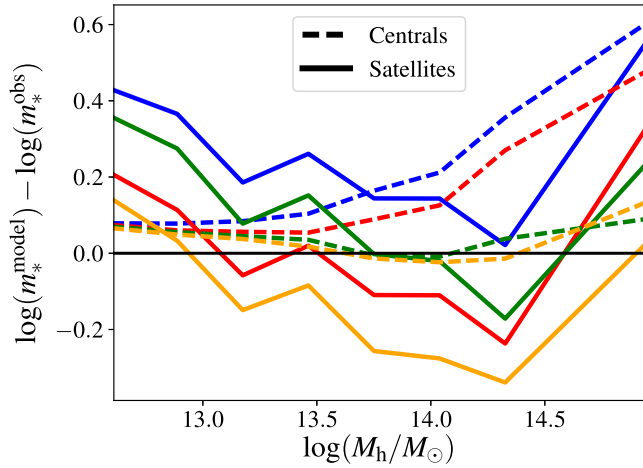
Figure 13. Differences between model predicted total stellar masses and observed total stellar masses as a function of host halo (group) mass, for central (*dashed curves*) and satellite (*solid curves*) galaxies. Models are colour-coded as in Figs 7, 10 and 12.
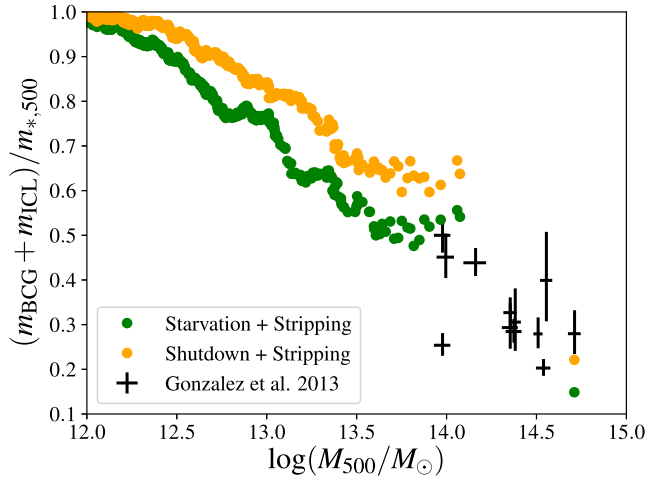
Figure 14. Fractional contribution of cD galaxy inclusive of its ICL to the total stellar mass of a group or a cluster. Only models with tidal stripping display an ICL component. Green and orange circles refer to the starvation model and the shutdown model, respectively. The crosses are the error bars for the observations of Gonzalez et al. (2013). Model predictions have been shown as a function of M_{500} (the total mass enclosed in sphere of radius R_{500}, within which the average density equals 500 times the critical density of the Universe) to match the definition of cluster mass used by Gonzalez et al. (2013).

standpoint. However, the Sérsic-exponential profile is nothing more than a useful fitting formula. Another functional form with more free parameters may fit the surface-brightness profile far beyond eight effective radii, eliminating the need for the ICL altogether. We do not question the claim by Bernardi et al. (2017) that PYMORPH returns objectively more accurate magnitudes than the SDSS pipeline. We enquire about the physical reason why giant ellipticals have extended light profiles, be they or not above a Sérsic-exponential fit. Following Gallagher & Ostriker (1972), we pursue the hypothesis that the extended envelopes of giant ellipticals are the debris of tidally disrupted galaxies, and define the ICL as the light from stars that have been tidally stripped from galaxies. This definition is of no assistance to an observer who wishes to measure the ICL. However, it is significant that when we compute the ICL mass according to
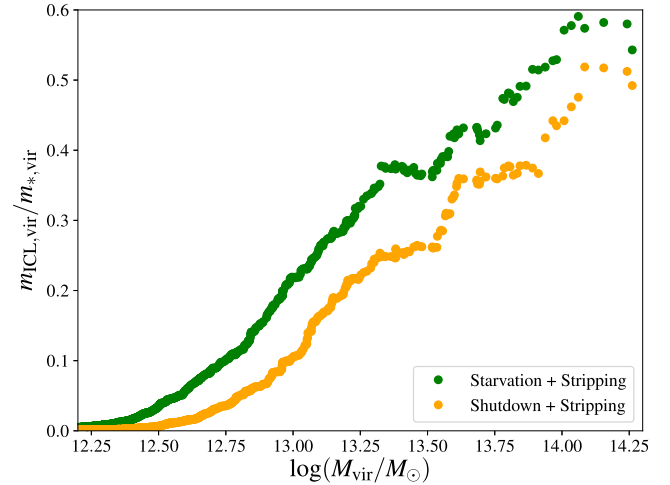
Figure 15. Fractional contribution of the outer envelopes to the total stellar masses within R_{500}. This represent the total diffuse light fraction within R_{vir}. The model with gas accretion (green) allows for more stripping than the model in which accretion shuts down immediately when a galaxy becomes a satellite (orange).

our definition, we recover a lot of the difference between the SMFs of Baldry et al. (2012) and (Bernardi et al. 2013, see Fig. 10, the gap between the blue and the green curves).

Bernardi et al. (2017) have also argued that the ICL should be centred on the centres of the clusters and should thus affect the magnitudes of central galaxies more than it affects those of satellites, while their work shows that, for a same luminosity, the difference between PYMORPH and SDSS magnitudes is about the same for both central and satellite galaxies. However, this is not a problem if one adopts our definition of the ICL because tidal stripping is expected to affect satellite galaxies, too. In fact, satellite galaxies will first develop long tidal tails and then these tails will coalesce into the extended envelopes of the central systems. This is how minor mergers have plausibly contributed to the considerable size evolution of elliptical galaxies from $z = 2$ to the present (e.g. Naab & Ostriker 2009; van Dokkum et al. 2010; Tal & van Dokkum 2011; Cooper et al. 2012; Shankar et al. 2013).

7 DISCUSSION

In this section, we discuss the uncertainties that affect our results. They come from (i) the resolution of the *N*-body simulation and the model for orphan galaxies that we have introduced to overcome the effects of limited resolution; (ii) uncertainties regarding the amount of star formation a satellite galaxy experiences post accretion on to the host halo; (iii) uncertainties regarding the amount of stellar mass loss experienced by satellite galaxies as a consequence of tidal stripping and heating; (iv) scatter in the SMHM relation; and (v) the AM method itself. We also discuss the excess of massive satellites in groups with $M_{\mathrm{h}} < 10^{13.44}$ M$_\odot$ predicted by all our models.

7.1 Model uncertainties

7.1.1 N-body resolution and orphan galaxies

In Section 3.2, we have treated ghost subhaloes as systems with well-defined orbits in a static spherical potential. Cosmological haloes contain substructures that contribute to their gravitational masses and perturb the orbital motions of subhaloes. Our work

does not consider the contribution of substructures to the gravitational potential of their host because our calculations are based on exclusive masses (our halo masses do not include the masses of substructures). We made this choice because the NFW model fits the density profiles of DM haloes more accurately when substructures are removed. If the mass distribution of substructures followed the NFW profile of the host halo, their merging time-scales would be shorter by typically 10 per cent. In reality, it is entirely possible that the interaction with other substructures may scatter a subhalo on an orbit with a longer merging time-scale. However, Hayashi, Navarro & Springel (2007) have shown that the isopotential surfaces inside a halo are much smoother than the density distribution and relatively insensitive to the presence of substructure.

The assumption that haloes are spherical is another simplification. Real haloes are triaxial. At $z = 0$, the typical minor-to-major axis ratio of the virial ellipsoid ranges from 0.75 at $M_h \sim 10^{11}$ M$_\odot$ to 0.5 at $M_h \sim 3 \times 10^{14}$ M$_\odot$ (Despali et al. 2017). Triaxiality increases at small radii but dissipation makes DM haloes substantially rounder at small radii than suggested by dissipationless simulations (Springel, White & Hernquist 2004). Furthermore, as expected from Poisson's equation, the gravitational potential tends to be much more spherical than the mass distribution. Indeed, Hayashi et al. (2007) find that a flattening (minor-to-major axis ratio) of ~0.4 in the mass distribution corresponds to a flattening of only ~0.75 for the isopotential contours the minor-to-major axis ratios of the isopotential contours are ≈0.75, hence much greater than the corresponding ratios for the density contours (≈0.4).

In this work, the approximation of a static spherical potential applies to ghost subhaloes only. At a given stellar mass, the fraction of satellite galaxies with unresolved (ghost) subhaloes depends on the resolution of the *N*-body simulation. If all subhaloes of satellite galaxies with $m_* > 10^9$ M$_\odot$ were resolved, our results would be independent of this approximation. Hence, while it is difficult to estimate, a priori, the errors introduced by treating ghost subhaloes as systems with well-defined orbits in a static spherical potential, it is easy to do it, a posteriori, by performing resolution studies.

To test the sensitivity of our results to *N*-body resolution and to our modelling of orphan galaxies, we have repeated our entire analysis on a simulation with the same cosmology, the same volume, the same initial conditions, but only 512^3 particles instead of 1024^3, and we allow ourselves to immediately merge orphan galaxies with the central galaxy of their host halo (i.e. to use the original merger tree without the addition of ghosts and orphan galaxies). We focus the comparison on our best-fitting model (starvation plus stripping) and on the mass range $10^{13.44}$ M$_\odot < M_h < 10^{13.73}$ M$_\odot$, but the results for this case also apply to the other models and mass bins.

Fig. 16 compares the conditional SMF for this model and mass range varying the resolution of the simulation and the treatment of orphans. With 512^3 particles, the conditional SMFs with (*solid purple curve*) and without (*dashed purple curve*) orphans differ at $m_* \lesssim 10^{10}$ M$_\odot$. However, with 1024^3 particles, the resolution is so good that delaying the mergers of orphans with central galaxies (*solid green curve*) or not (*dashed green curve*) makes little difference above $m_* = 10^{8.5}$ M$_\odot$. The treatment of orphans is a small correction and therefore a negligible source of uncertainty in relation to our conclusions.

Above $m_* \sim 10^{10}$ M$_\odot$, the conditional SMFs for the 512^3 simulation without orphans (*dashed purple curve*) and the 1024^3 simulation without orphans (*dashed green curve*) are very similar, suggesting that numerical convergence has been reached. Most interesting, however, is the agreement of the 512^3 and 1024^3 simulations when orphans are included, as we see convergence (solid green and purple
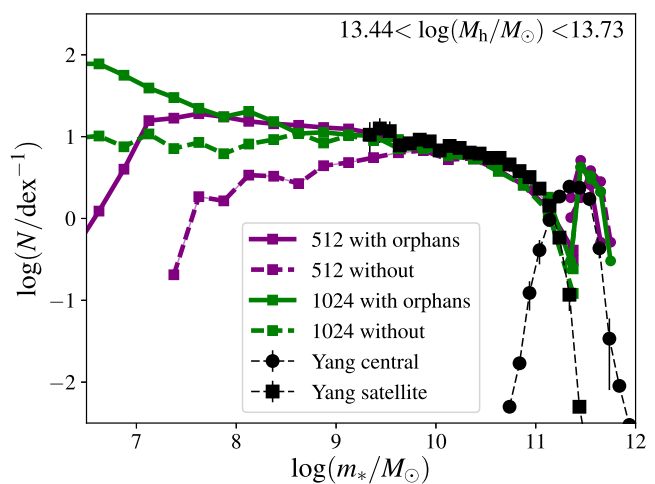
Figure 16. Sensitivity of the conditional SMF predicted by the starvation+stripping model to the resolution of the *N*-body simulation, degraded from 1024^3 (*green*) to 512^3 particles (*purple lines*), and to the merger time of orphan galaxies with the central galaxy: immediately (*dashed lines*) or only at the first pericentre following the expected time of orbital decay by dynamical friction (*solid lines*). The figure shows the case for haloes with $13.44 < \log(M_h/M_\odot) < 13.73$. In the mass range probed by the observations (Yang et al. 2012; *black symbols* with error bars), the model with orphan galaxies has converged because the simulations with 512^3 and 1024^3 particles give very similar conditional SMFs.

lines) in the conditional SMF down to $10^{7.7}$ M$_\odot$. This proves that the inclusion of orphans aids in achieving convergence to correct solution (also see Guo et al. 2011). At $m_* \gtrsim 3 \times 10^8$ M$_\odot$, the 512^3 simulation with orphans is at least as good as the 1024^3 simulation without orphans.

7.1.2 Gas accretion on to satellites

Estimating how much gas accretes on to satellite galaxies after entering a group or cluster environment is less straightforward than testing for resolution effects, but the simple assumption that star formation shuts down immediately cannot be correct. Weinmann et al. (2006) used a version of the Munich SAM in which there was no accretion on to satellite galaxies (Croton et al. 2006). They found that the fraction of faint satellites with red colours was overestimated by a factor of ~2–3. All SAMs published in those years shared the same problem (e.g. Fontanot et al. 2009). Indeed, Cattaneo et al. (2007) ran the GalICS SAM on merger trees from a cosmological hydrodynamic simulation. GalICS assumed no gas accretion on satellites and predicted a much higher fraction of quenched galaxies than the hydrodynamic simulation.

A delayed quenching scenario can be parametrized by two time-scales: the time t_{delay} during which a galaxy keeps forming stars after entering a group or a cluster and the t_{quench}, over which the star formation rate rapidly decays after t_{delay} has elapsed. Several authors have investigated these time-scales. Mahajan, Mamon & Raychaudhury (2011) split galaxies between infalling, backsplash and virialized, and combined the fraction of star-forming galaxies observed in the SDSS with cosmological *N*-body simulations to quantify projection effects. Their analysis suggests that quenching is delayed until galaxies reach the virial radius on their way out of the cluster after the first pericentric passage. As shown in Fig. B1 in Appendix B, galaxies that are at the virial radius today, on their way out after their first pericentric passage, and have typical first apocentres close to the turnaround radius at that time (3–4 virial

radii at that time), entered the group/cluster environment ∼3 Gyr ago (Fig. B1 in Appendix B), and passed through the pericentre ∼1.6 Gyr ago. Therefore, according to the modelling of Mahajan et al., star formation is quenched ∼3 Gyr after cluster entry and ∼1.6 Gyr after the first pericentric passage. Wetzel et al. (2013) used an *N*-body simulation to measure the characteristic time, since t_{entry} of a galaxy at a given R/R_{vir}, and constrained t_{delay} and t_{quench} by measuring the fraction of red galaxies in SDSS groups/clusters as a function of the distance from the centre. A slow progressive fading of star formation since t_{entry} would blur the bimodal distribution of galaxy colour. In contrast, the observations are consistent with a long delay ($t_{delay} = 2$–4 Gyr) followed by rapid quenching ($t_{quench} = 0.2$–0.8 Gyr). Haines et al. (2015) performed a similar study to match the observed distribution of the fraction of star-forming galaxies with the predictions of times since entry as a function of position in projected phase space from cosmological *N*-body simulations. They conclude that star formation declines exponentially after entering the virial radius on a time-scale of 1.7 Gyr. A similar analysis by Oman & Hudson (2016) suggests that star formation in cluster satellite galaxies is rapidly quenched within ∼1–2 Gyr from the first pericentric passage. Another recent similar study based on both SDSS and higher redshift data leads to delay times of 2–5 Gyr (Fossati et al. 2017).

Further evidence in support of delayed quenching comes from chemical abundances. When a galaxy ceases to accrete pristine gas but keeps forming stars, its metal content relative to hydrogen increases. From the metal abundances of red galaxies with $m_* \lesssim 10^{10.5}$ M$_\odot$, Peng et al. (2015) inferred that they must have behaved as closed boxes for ∼4 Gyr before they eventually run out of gas. The higher metallicities of satellite galaxies were interpreted as evidence that this is due to starvation by the environment. While the complete starvation of gas accretion in Peng et al.'s picture seems to conflict with SAMs.[10] There is consensus that star formation cannot have been quenched instantaneously at t_{entry}.

In conclusion, while it is not straightforward to determine what fraction of the gas associated with a subhalo will accrete on to the satellite galaxy it contains and what fraction will be stripped (mainly by ram pressure, which is more important than tidal stripping for gas[11]), and while the precise value will also depend on the feedback one assumes (Tomozeiu, Mayer & Quinn 2016), there appears to be observational consensus that star formation is quenched 1–2 Gyr after the first pericentric passage. Therefore, the starvation model, which prevents further accretion from the environment but allows star formation to continue until the first pericentric passage, seems a much more plausible assumption than to assume a complete shutdown of star formation at the entry time.

7.1.3 Tidal stripping

Tidal stripping is an inevitable dynamical process, but its analytic modelling is not straightforward and requires simplifying assumptions. The most common assumption is instantaneous tides applied

---

[10] The difference is largely due to the assumed star formation efficiencies. SAMs usually assume shorter star formation time-scales than those of Peng et al. (2015), at least at low stellar mass. Hence, they need sustained accretion to keep star formation going for several Gyr.

[11] In field galaxies, HI discs are more extended than stellar discs. In satellite galaxies, it is often the contrary because ram pressure has stripped their outer parts. Where tidal stripping the dominant phenomenon, the HI disc would be truncated at the same radius as the stellar disc because tides do not differentiate between gas and stars.
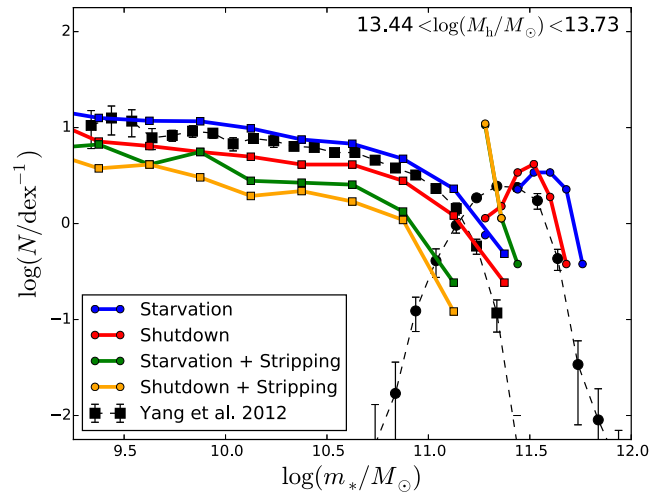
Figure 17. Conditional SMF for our different models (curves) compared to the observations of Yang et al. (2012; black data points with error bars). This figure is identical to the $10^{13.44}$ M$_\odot$ $< M_h < 10^{13.73}$ M$_\odot$ panel of Fig. 12 except for the calculation of tidal stripping (the green and orange curves). Here, tidal stripping is computed in the circular-orbit instantaneous-tide approximation, using equation (3) with $\alpha = -3$.

to satellites are on circular orbits. However, for a fixed pericentric radius, this assumption gives an upper limit rather than a realistic estimate for the stellar mass that is tidally stripped from galaxies because most orbits are highly elongated. The circular-orbit approximation underpredicts the conditional SMF even when it is applied to the starvation model, which corresponds to the maximum possible star formation in satellite galaxies (*green curve* in Fig. 17) because it results in stellar mass loss of satellite galaxies that can be as large as 0.5 dex. On the contrary, the difference between the starvation model without stripping (*blue curve*) and the observations (*horizontal black line*) in Fig. 13 shows that the stellar mass that can plausibly be stripped from galaxies is $\lesssim$0.15–0.2 dex ($\lesssim$0.1–0.15 dex if we allow for $\gtrsim$10 per cent stellar mass loss through stellar evolution between z_{entry} and $z = 0$). This upper limit is obtained by comparing a model without stripping to the observations. It is therefore totally independent of any physical model of tidal stripping.

Re-assuringly, the more sophisticated model in Section 5.2 predicts tidal stripping by ∼0.1 dex (∼25 per cent) on average, in agreement with the upper limit above. This value (based on the green curve in Fig. 13) has been computed assuming the maximum tidal acceleration but also assuming that the tidal acceleration acts only for a very short time around the pericentric passage (corresponding to the part of the orbit shown as a *thick* solid red line in Fig. 1). Using the average acceleration for a test particle (equal to half the maximum acceleration; Appendix A) while retaining the second assumption will most likely underestimate the tides. However, this results in an average tidal stripping of 0.07–0.08 dex on average, so the quantitative difference is small.

The assumption of circular orbits is as incorrect for the DM as it is for the stars, but we kept using it to compute the DM lost by ghost subhaloes because DM is stripped all the way down to the centre, not just at pericentre (Klimentowski et al. 2009), so the impulsive approximation is not necessarily much more accurate. The question is the extent to which its inaccuracy affects our conclusions. Tidal radii have no consequences on the survival times of ghost subhaloes, which are by Jiang et al.'s (2008) formula. Their only effect on the

tidal stripping of stars is through the value of R_p. If a ghost subhalo is stripped too heavily (the most likely outcome of our approximation), it will suffer less dynamical friction. Less dynamical friction implies less orbital decay. The pericentric radius will be overestimated and the tidal stripping of stars will be less efficient than for the correct value of R_p. Therefore, we can be confident that 0.07–0.08 dex is a plausible lower limit for the stellar mass lost by galaxies owing to dynamical friction.

In practice, ghost subhaloes (and thus this calculation) were introduced as a way to beat the resolution limit of our *N*-body simulation. A posteriori, our resolution is so good that conditional SMFs are very similar with or without them (Fig. 16). Thus, any error in our calculation of the pericentric radii of ghost subhaloes is bound to have a limited impact on the conclusions of this article.

7.1.4 Scatter in the SMHM relation

Our AM procedure (Section 1.2) assumes that the m_*–M_{max} relation does not contain any scatter. In this section, we discuss how scatter can affect our conclusions.

Observationally, to determine the scatter in stellar mass at constant halo mass, one needs a method to measure M_h. Yang et al. (2009) estimated the masses of groups from their luminosities (Section 6.2) and found a scatter in $\log m_*$ of $\sigma_{\log m_*} \simeq 0.17$ dex in m_* at constant M_h. More et al. (2009) found a similar result ($\sigma_{\log m_*} \simeq 0.16$ dex) using halo masses from satellite kinematics. Leauthaud et al. (2012) performed a more sophisticated analysis by fitting simultaneously the galaxy SMF, clustering data (correlation functions) and halo masses from galaxy–galaxy lensing. They found an intrinsic scatter of about 0.2 dex after subtracting errors from photometry, photometric redshifts and spectral energy distribution fitting. Behroozi et al. (2013) used AM to infer a scatter of 0.22 ± 0.02 dex. Coupon et al. (2015) repeated the same analysis with more recent data and confirmed their results.

Implementing scatter in our models requires a more sophisticated approach than simply applying random errors to the stellar masses determined from the AM relation. If we simply perturbed the AM relation, our models would no longer reproduce the galaxy SMF because of the Eddington bias. We overcome this problem by splitting the galaxy population into pairs. One galaxy has logarithmic stellar mass $\log m_*$. The other has logarithmic stellar mass $\log m_* + \log \Delta m_*$, where $\log \Delta m_*$ is a random number from a Gaussian distribution with standard deviation $\sigma_{\log m_*}$ and zero mean. Scatter is implemented by swapping the haloes of the two galaxies. This swapping introduces the requires scatter in the SMHM relation without changing the actual SMF of the galaxies.

This procedure means that the conditional SMFs computed by our models are now dependent on the random way in which galaxy population has been splitted into pairs but we can obtain robust results by averaging over many different realizations. The thin green dashed curves in Fig. 12 show the median conditional SMF for the starvation plus stripping model over a hundred realizations with $\sigma_{\log m_*} = 0.2$, while the green shaded areas show upper and lower quartiles for the same hundred realizations. The absence of systematic differences between the thin and the thick green curves (the model without scatter) proves that scatter adds noise but will not bias our conclusion. Thin lines and shaded areas have been shown for the starvation plus stripping model only not to overcrowd the figure.

7.1.5 Uncertainties in the SMHM relation

Fig. 6 shows that the SMHM relations derived with different AM/HOD models differ at the level of 0.1–0.2 dex in stellar mass. One could interpret these differences as a measure of the intrinsic uncertainty of the SMHM relation from AM. To understand the implications that such an uncertainty may have for our results, we begin by discussing the origin of these differences.

The first source of difference is the SMF used to constrain the SMHM relation. The SMHM relation of Moster et al. (2013) differs from those of other authors at low masses because they used the SMF of Li & White (2009), which contains a higher a number density of galaxies with $10^9 \, M_\odot < m_* < 10^{10.2} \, M_\odot$ than the SMFs of Baldry, Glazebrook & Driver (2008), Baldry et al. (2012), Leauthaud et al. (2012), Yang et al. (2012), Papastergis et al. (2012) and Moustakas et al. (2013).

Secondly, the HMFs assumed by different authors can come either from *N*-body simulations or from the Sheth & Tormen (2002) formula, which is calibrated on *N*-body simulations. Even if the cosmologies assumed by different authors were completely identical, there would still be an uncertainty of about 10 per cent in the HMF from the halo finder (Knebe et al. 2013).

Finally, one can use the AM method, as we have done, or one can assumed a parametric SMHM relation and constrain its parameters so that it fits the SMF. The results obtained with the two methods will be very similar but not necessarily identical. One can also consider or not consider the presence of scatter (Section 7.1.4). Models with scatter find lower stellar masses for a given halo mass at high masses to compensate for the Eddington bias.

We argue that these systematic uncertainties are not important for our conclusions because our analysis focuses on the differences in stellar mass between field and satellite galaxies. We apply the same *N*-body simulation, the same halo finder and the same AM procedure to both. Therefore, these uncertainties cancel out in the relative comparison, as would any systematic error in the photometry or the initial mass function.

The only real question is whether our SMF $n_*(m_*, z)$, which is constructed from data at different redshifts (Yang et al. 2012 at $z < 0.2$, Muzzin et al. 2013 at $0.2 < z < 2.5$), is fully consistent with the conditional SMF of Yang et al. (2012), to which we compare our results. Fig. 5 shows that, for $10^9 \, M_\odot < m_* < 10^{11.5} \, M_\odot$ the local SMF assumed for this work (black solid curve) is fully consistent with both the SMF of Yang et al. (2012; black symbols with error bars) and the SMF of Muzzin et al. (2013) extrapolated to $z = 0.1$ (black dashed curve).

7.2 The massive satellite excess in low-mass groups

The most noticeable discrepancy between our models and the observations of Yang et al. (2012) is the excess of massive satellites with $m_* \gtrsim 10^{11} \, M_\odot$ in low-mass groups (Fig. 12). This excess cannot be due to star formation after z_{entry} or to underestimated stripping because, at $M_h \lesssim 10^{13} \, M_\odot$ it is present even in the shutdown model when tidal stripping is computed with the instantaneous-tide circular-orbit approximation (which largely overestimates the magnitude of the phenomenon; Section 7.1.1).

There are two possible explanations for this discrepancy. First, Yang et al. (2012) may have classified as central systems that, in the *N*-body simulation, our halo finder classifies as satellites, see Skibba et al. (2011) and Lange et al. (2017) for a discussion of this phenomena. Bernardi et al. (2017) have analysed SDSS groups with a group finder called redMaPPer, which differ from the one

used by Yang et al. (2012). They have remarked that: 'Many of the objects which Yang et al. classify as being centrals in groups less massive than $10^{14}\,\mathrm{M}_\odot$ are called satellites by REDMAPPER'. Second, a satellite with m_* comparable to the central galaxy in a small group of often 2–3 objects is very different system from a satellite in a cluster, whose total mass is much larger than that of any satellite. The Jiang et al. (2008) formula (combined with our prescription, which requires one galaxy out of four to complete an additional orbit after a time t_{df} has elapsed since entry) may fail when applied to nearly equal-mass binary systems. If their merging time is systematically overestimated, this could explain the excess of massive satellites in low-mass groups.

8 CONCLUSIONS

The original goal of this work was to estimate the stellar mass lost by galaxies in groups and clustering due to tidal stripping by comparing the distribution of entry masses to observations of the conditional SMF. As our work progressed, we realized that this original approach was too simplistic because the distribution of entry masses (shown by the red curves in Fig. 12, except for satellites that have merged) is below the data points for most values of m_*. If we look at the mass variation from z_{entry} to $z = 0$, then galaxies have gained stellar mass, not lost it.

The simplest refinement of this analysis is to assume that $\Delta m_{strip} = m_{entry} + \Delta m_* - m_*$, where m_{entry} is the stellar mass at z_{entry}, Δm_* is the mass of the stars formed between z_{entry} and $z = 0$ and m_* is the stellar mass at $z = 0$. The problem is that Δm_* is considerably uncertain. We can obtain an upper limit for Δm_{strip} by assuming that, in satellite galaxies, stellar mass grows with halo mass following the same relation that holds for central galaxies. This assumption maximizes Δm_*. If we follow this approach, we find that the stellar mass loss from satellite galaxies is $\lesssim 0.15$–0.2 dex. However, this would imply that central and satellite galaxies have similar SFRs. This is inconsistent with observations, which indicate that satellite galaxies have, on average, lower SFRs than centrals of the same stellar mass (i.e. Weinmann et al. 2006; Wetzel et al. 2013, and references therein). Hence, this model clearly has to be regarded an extreme upper limit for Δm_*.

The upper limit for Δm_{strip} includes both the stellar mass lost due to tidal stripping and the decrease in stellar mass that results from stellar evolution. A typical satellite galaxy is accreted into its host halo at a median redshift of $z_{entry} \sim 0.5$, which corresponds to a lookback time of $\sim$5 Gyr. Assuming passive evolution, a typical quiescent galaxy will lose between 10 and 20 per cent of its mass over a period of $\sim$5 Gyr (e.g. Fioc & Rocca-Volmerange 1997). Accounting for this passive evolution, gives a more stringent upper limit $\Delta m_{strip} \lesssim 0.1$–$0.15$ dex.

This is an upper limit because of the uncertainty on Δm_*. Since $m_{entry} - m_* < 0$ and $m_{entry} + \Delta m_*^{max} - m_* > 0$, it is possible to find a plausible value of Δm_* for which $\Delta m_{strip} = m_{entry} + \Delta m_* - m_* = 0$, that is, the analysis above cannot rule out a model without stripping.

We have compared this indirect result with direct analytic estimates of the stellar masses that galaxies lose due to tidal stripping. The simplest estimates based on instantaneous tides and circular orbits are highly inaccurate because most satellites are on highly elongated orbits. These estimates predict much more stripping that is allowed by the upper limit derived in this article. More sophisticated estimates assume impulsive stripping on elongated orbits (satellite galaxies lose stars at each pericentric passage). In this article, we have improved previous analytic models of impulsive

tides (Spitzer 1958; Gonzalez-Casado et al. 1994; Mamon 2000) and have used our results (equation 18) to predict stripping by 0.07–0.1 dex on average, which is consistent with our upper limit $\lesssim 0.1$–0.15 dex.

We consider 0.07 dex (17 per cent) to be a reasonable lower limit for the stellar mass lost by galaxies owing to tidal stripping because: (i) its calculation is based on the average rather than the maximum tidal acceleration; (ii) we assumed that the tides acts only for a very short time interval around the pericentre and (iii) pericentric radii that may be overestimated (in case of ghost subhaloes) but not underestimated (tides are stronger for closer pericentric passages).

Our best estimate for the stellar mass lost owing to tidal stripping, $\sim$0.07–0.1 dex ($\sim$17–25 per cent), is consistent with a picture in which Δm_* is close to Δm_*^{max}, i.e. one in which satellite galaxies are quenched several Gyr after entering a group or cluster environment and, in any case, after the first pericentric passage (Mahajan et al. 2011; Wetzel et al. 2013; Haines et al. 2015; Peng et al. 2015; Oman & Hudson 2016; Fossati et al. 2017).

Our model predicts that the fraction of stars that contribute the ICL increase with the mass of the host system. In clusters, stars tidally stripped from galaxies are predicted to contribute to half of the total light within R_{500}.

ACKNOWLEDGEMENTS

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., 2006, MNRAS, 365, 747
Aubert D., Pichon C., Colombi S., 2004, MNRAS, 352, 376
Avila S. et al., 2014, MNRAS, 441, 3488
Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, ApJ, 600, 681
Baldry I. K., Glazebrook K., Driver S. P., 2008, MNRAS, 388, 945
Baldry I. K. et al., 2012, MNRAS, 421, 621
Bautz L. P., Morgan W. W., 1970, ApJ, 162, L149
Behroozi P. S., Conroy C., Wechsler R. H., 2010, ApJ, 717, 379
Behroozi P. S., Wechsler R. H., Conroy C., 2013, ApJ, 770, 57
Bekki K., 2009, MNRAS, 399, 2221
Bell E. F. et al., 2008, ApJ, 680, 295
Benson A. J., 2012, New Astron., 17, 175
Bernardi M., Roche N., Shankar F., Sheth R. K., 2011a, MNRAS, 412, 684
Bernardi M., Roche N., Shankar F., Sheth R. K., 2011b, MNRAS, 412, L6
Bernardi M., Meert A., Sheth R. K., Vikram V., Huertas-Company M., Mei S., Shankar F., 2013, MNRAS, 436, 697
Bernardi M., Fischer J.-L., Sheth R. K., Meert A., Huertas-Company M., Shankar F., Vikram V., 2017, MNRAS, 468, 2569
Binney J., Tremaine S., 2008, Galactic Dynamics: Second Edition. Princeton Univ. Press, Princeton, NJ
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, MNRAS, 370, 645
Boylan-Kolchin M., Ma C.-P., Quataert E., 2008, MNRAS, 383, 93
Bryan G. L., Norman M. L., 1998, ApJ, 495, 80
Bullock J. S., Dekel A., Kolatt T. S., Kravtsov A. V., Klypin A. A., Porciani C., Primack J. R., 2001, ApJ, 555, 240
Cattaneo A., Dekel A., Devriendt J., Guiderdoni B., Blaizot J., 2006, MNRAS, 370, 1651
Cattaneo A. et al., 2007, MNRAS, 377, 63

Cattaneo A., Mamon G. A., Warnick K., Knebe A., 2011, A&A, 533, A5
Chan R., Mamon G. A., Gerbal D., 1997, Astrophys. Lett. Commun., 36, 47
Chandrasekhar S., 1943, ApJ, 97, 255
Cole S., Lacey C., 1996, MNRAS, 281, 716
Conroy C., Wechsler R. H., Kravtsov A. V., 2006, ApJ, 647, 201
Conroy C., Wechsler R. H., Kravtsov A. V., 2007, ApJ, 668, 826
Copper M. C. et al., 2012, MNRAS, 419, 3018
Cora S. A., 2006, MNRAS, 368, 1540
Cora S. A., Muzzio J. C., Vergne M. M., 1997, MNRAS, 289, 253
Coupon J. et al., 2015, MNRAS, 449, 1352
Croton D. J. et al., 2006, MNRAS, 365, 11
de la Torre S. et al., 2011, A&A, 525, A125
De Lucia G., Blaizot J., 2007, MNRAS, 375, 2
Dekel A., Birnboim Y., 2006, MNRAS, 368, 2
Dekel A., Devor J., Hetzroni G., 2003, MNRAS, 341, 326
Despali G., Giocoli C., Bonamigo M., Limousin M., Tormen G., 2017, MNRAS, 466, 181
Duarte M., Mamon G. A., 2015, MNRAS, 453, 3848
Ferguson A. M. N., Irwin M. J., Ibata R. A., Lewis G. F., Tanvir N. R., 2002, AJ, 124, 1452
Fioc M., Rocca-Volmerange B., 1997, A&A, 326, 950
Fontanot F., Somerville R. S., Silva L., Monaco P., Skibba R., 2009, MN-RAS, 392, 553
Fossati M. et al., 2017, ApJ, 835, 153
Gallagher J. S., III, Ostriker J. P., 1972, AJ, 77, 288
Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, MNRAS, 362, 41
Gargiulo I. D. et al., 2015, MNRAS, 446, 3820
Ghigna S., Moore B., Governato F., Lake G., Quinn T., Stadel J., 1998, MNRAS, 300, 146
Gonzalez A. H., Zabludoff A. I., Zaritsky D., 2005, ApJ, 618, 195
Gonzalez A. H., Sivanandam S., Zabludoff A. I., Zaritsky D., 2013, ApJ, 778, 14
Gonzalez-Casado G., Mamon G. A., Salvador-Sole E., 1994, ApJ, 433, L61
Gonzalez-Perez V., Lacey C. G., Baugh C. M., Lagos C. D. P., Helly J., Campbell D. J. R., Mitchell P. D., 2014, MNRAS, 439, 264
Guo Q., White S., Li C., Boylan-Kolchin M., 2010, MNRAS, 404, 1111
Guo Q. et al., 2011, MNRAS, 413, 101
Haines C. P. et al., 2015, ApJ, 806, 101
Hatton S., Devriendt J. E. G., Ninin S., Bouchet F. R., Guiderdoni B., Vibert D., 2003, MNRAS, 343, 75
Hayashi E., Navarro J. F., Taylor J. E., Stadel J., Quinn T., 2003, ApJ, 584, 541
Hayashi E., Navarro J. F., Springel V., 2007, MNRAS, 377, 50
Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, Nature, 402, 53
Henriques B. M. B., Thomas P. A., 2010, MNRAS, 403, 768
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194
Jiang C. Y., Jing Y. P., Faltenbacher A., Lin W. P., Li C., 2008, ApJ, 675, 1095
Kang X., van den Bosch F. C., 2008, ApJ, 676, L101
Kang X., Jing Y. P., Mo H. J., Börner G., 2005, ApJ, 631, 21
Kauffmann G. et al., 2003, MNRAS, 346, 1055
Kazantzidis S., Łokas E. L., Callegari S., Mayer L., Moustakas L. A., 2011, ApJ, 726, 98
Kazantzidis S., Łokas E. L., Mayer L., 2013, ApJ, 764, L29
Klimentowski J., Łokas E. L., Kazantzidis S., Mayer L., Mamon G. A., 2009, MNRAS, 397, 2015
Knebe A. et al., 2011, MNRAS, 415, 2293
Knebe A. et al., 2013, MNRAS, 435, 1618
Knebe A. et al., 2015, MNRAS, 451, 4029
Krick J. E., Bernstein R. A., Pimbblet K. A., 2006, AJ, 131, 168
Lange J. U., van den Bosch F. C., Hearin A., Campbell D., Zentner A. R., Villarreal A., Mao Y.-Y., 2017, preprint (arXiv:1705.05043)
Leauthaud A. et al., 2010, ApJ, 709, 97
Leauthaud A. et al., 2012, ApJ, 744, 159
Lee J., Yi S. K., 2013, ApJ, 766, 38
Li C., White S. D. M., 2009, MNRAS, 398, 2177
Lin Y.-T., Mohr J. J., 2004, ApJ, 617, 879

Liu L., Yang X., Mo H. J., van den Bosch F. C., Springel V., 2010, ApJ, 712, 734
Łokas E. L., Kazantzidis S., Mayer L., 2011, ApJ, 739, 46
Mahajan S., Mamon G. A., Raychaudhury S., 2011, MNRAS, 416, 2882
Mamon G. A., 1987, ApJ, 321, 622
Mamon G., 1996, in de Vega H. J., Sánchez N., eds, Third Paris Cosmology Colloquium, World Scientific Press, Singapore, p. 95
Mamon G. A., 2000, in Combes F., Mamon G. A., Charmandaris V., eds, ASP Conf. Ser. Vol. 197, Dynamics of Galaxies: from the Early Universe to the Present. Astron. Soc. Pac., San Francisco, p. 377
Mamon G. A., Łokas E. L., 2005, MNRAS, 363, 705
Marinoni C., Hudson M. J., 2002, ApJ, 569, 101
Matthews T. A., Morgan W. W., Schmidt M., 1964, ApJ, 140, 35
McCarthy I. G., Frenk C. S., Font A. S., Lacey C. G., Bower R. G., Mitchell N. L., Balogh M. L., Theuns T., 2008, MNRAS, 383, 593
Merritt D., 1983, ApJ, 264, 24
Mihos J. C., Harding P., Feldmeier J., Morrison H., 2005, ApJ, 631, L41
Mo H. J., Mao S., White S. D. M., 1998, MNRAS, 295, 319
More S., van den Bosch F. C., Cacciato M., Mo H. J., Yang X., Li R., 2009, MNRAS, 392, 801
Morgan W. W., Lesh J. R., 1965, ApJ, 142, 1364
Moster B. P., Naab T., White S. D. M., 2013, MNRAS, 428, 3121
Moustakas J. et al., 2013, ApJ, 767, 50
Muzzin A. et al., 2013, ApJ, 777, 18
Naab T., Ostriker J. P., 2009, ApJ, 690, 1452
Navarro J. F., Frenk C. S., White S. D. M., 1996, ApJ, 462, 563
Odenkirchen M., Grebel E. K., Dehnen W., Rix H.-W., Cudworth K. M., 2002, AJ, 124, 1497
Oman K. A., Hudson M. J., 2016, MNRAS, 463, 3083
Ostriker J. P., Spitzer L., Jr, Chevalier R. A., 1972, ApJ, 176, L51
Papastergis E., Cattaneo A., Huang S., Giovanelli R., Haynes M. P., 2012, ApJ, 759, 138
Pasquali A., Gallazzi A., Fontanot F., van den Bosch F. C., De Lucia G., Mo H. J., Yang X., 2010, MNRAS, 407, 937
Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192
Planck Collaboration XX, 2014, A&A, 571, A20
Pranger F., Trujillo I., Kelvin L. S., Cebrián M., 2017, MNRAS, 467, 2127
Press W. H., Schechter P., 1974, ApJ, 187, 425
Prugniel P., Combes F., 1992, A&A, 259, 25
Pujol A. et al., 2017, MNRAS, 469, 749
Purcell C. W., Bullock J. S., Zentner A. R., 2007, ApJ, 666, 20
Reyes R., Mandelbaum R., Gunn J. E., Nakajima R., Seljak U., Hirata C. M., 2012, MNRAS, 425, 2610
Shankar F., Marulli F., Bernardi M., Mei S., Meert A., Vikram V., 2013, MNRAS, 428, 109
Sheth R. K., Tormen G., 2002, MNRAS, 329, 61
Skibba R. A., van den Bosch F. C., Yang X., More S., Mo H., Fontanot F., 2011, MNRAS, 410, 417
Somerville R. S., Primack J. R., 1999, MNRAS, 310, 1087
Spitzer Jr L., 1958, ApJ, 127, 17
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Springel V., White S. D. M., Hernquist L., 2004, in Ryder S., Pisano D., Walker M., Freeman K., eds, IAU Symp. 220, Dark Matter in Galaxies, The Shapes of Simulated Dark Matter Halos. Astron. Soc. Pac., San Francisco, p. 421
Srisawat C. et al., 2013, MNRAS, 436, 150
Tal T., van Dokkum P. G., 2011, ApJ, 731, 89
Tomozeiu M., Mayer L., Quinn T., 2016, ApJ, 827, L15
Tweed D., Devriendt J., Blaizot J., Colombi S., Slyz A., 2009, A&A, 506, 647
Vale A., Ostriker J. P., 2004, MNRAS, 353, 189
van den Bosch F. C., Jiang F., Hearin A., Campbell D., Watson D., Padman-abhan N., 2014, MNRAS, 445, 1713
van den Bosch F. C., Jiang F., Campbell D., Behroozi P., 2016, MNRAS, 455, 158
van Dokkum P. G. et al., 2010, ApJ, 709, 1018
Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., Croton D. J., Moore B., 2006, MNRAS, 372, 1161

Welch G. A., Sastry G. N., 1971, ApJ, 169, L3

Wetzel A. R., Tinker J. L., Conroy C., van den Bosch F. C., 2013, MNRAS, 432, 336

Willman B., Governato F., Wadsley J., Quinn T., 2004, MNRAS, 355, 159

Yang X., Mo H. J., van den Bosch F. C., 2009, ApJ, 693, 830

Yang X., Mo H. J., van den Bosch F. C., Zhang Y., Han J., 2012, ApJ, 752, 41

Yanny B. et al., 2003, ApJ, 588, 824

Yi S. K., Lee J., Jung I., Ji I., Sheen Y.-K., 2013, A&A, 554, A122

Zibetti S., White S. D. M., Schneider D. P., Brinkmann J., 2005, MNRAS, 358, 949

Zwicky F., 1957, Morphological Astronomy. Springer, Berlin

APPENDIX A: THEORY OF TIDES IN THE CIRCULAR-ORBIT APPROXIMATION

In this appendix, we compute the tidal acceleration $\boldsymbol{a}_\text{t}$ that a host system exerts on a test particle of a satellite. We also compute the tidal radius r_t of the satellite in the approximations that: (i) the satellite is on a circular orbit, so that the gravitational potential is static in a corotating frame, and (ii) the test particle is instantaneously stripped as soon as $\boldsymbol{a}_\text{t}$ exceeds the gravitational acceleration that keeps the particle bound to the satellite. The latter is an approximation because a net outward acceleration is a necessary but not sufficient condition for tidal stripping. For it to be sufficient, the speed $\int \boldsymbol{a}_\text{t}\,\text{d}t$ imparted by the acceleration to the particle must be large enough to unbind it. In Section 5.2, we build on these results and generalize them to non-circular orbits and non-instantaneous tides.

Let M_h and M_s be the masses of the host (of centre of mass H) and the satellite (of centre of mass S). Let H and S be the respective centres of mass of the halo and subhalo, call O the centre of mass of the halo+subhalo system and P the position of a particle in the subhalo. Finally, denote $\boldsymbol{r} \equiv \vec{\text{SP}}$ and $\boldsymbol{R} \equiv \vec{\text{HS}}$. Because of assumption (i), H and S rotate around O with angular velocity $\boldsymbol{\Omega}$, such that

$$\Omega = \sqrt{\frac{G\,(M_\text{h} + M_\text{s})}{R^3}}\,. \tag{A1}$$

In a corotating reference frame, the satellite is subject to two accelerations that cancel one another: the gravitational attraction of the central system and the centrifugal acceleration. The particle P is subject to four accelerations: the gravitational attraction of the satellite, the gravitational attraction of the host, the centrifugal acceleration and the Coriolis acceleration. The sum of the last three accelerations defines the tidal acceleration $\boldsymbol{a}_\text{t}$, which can be written

$$\dot{\boldsymbol{v}} = \boldsymbol{a}_\text{t} - \nabla\Phi_\text{s}\,, \tag{A2}$$

$$\boldsymbol{a}_\text{t} = -\nabla\Phi_\text{h} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \vec{\text{OP}}) - 2\boldsymbol{\Omega} \times \boldsymbol{v}\,, \tag{A3}$$

where $\boldsymbol{v}$ is the velocity of the particle in the corotating frame, Φ_h is the gravitational potential of the host and Φ_s is the gravitational potential of the satellite. Assumption (ii) applied to equation (A3) implies that the particle will be tidally stripped when $|\boldsymbol{a}_\text{t}| > \nabla\Phi_\text{s}$ in equation (A3).

To proceed further, we must make additional assumptions. Here, we assume that: (iii) P is in the orbital plane of the binary, so that $-\boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \vec{\text{OP}}) = \Omega^2\vec{\text{OP}}$ and (iv) P turns around S on a circular orbit with angular velocity $\boldsymbol{\Omega}$, so that $\boldsymbol{v} = 0$ in the corotating frame (as it is the case for the Moon, which corotates in phase locking with the Earth). The impact of these assumptions on the value of r_t will
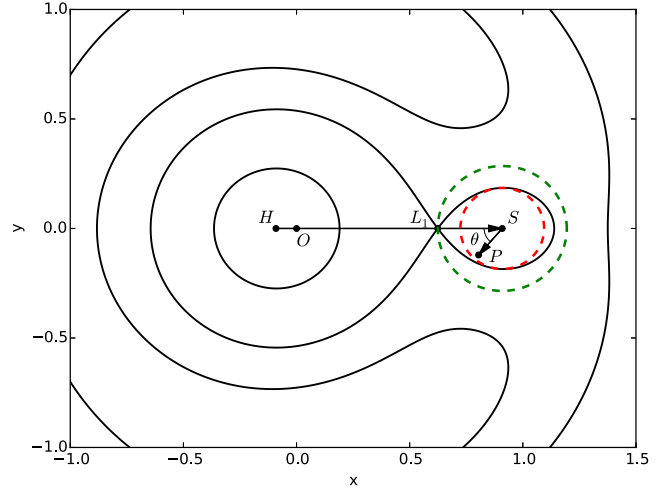
Figure A1. Equipotential curves for the effective potential Φ_eff (equation A4) of a two-body system composed of a host halo and a subhalo, centred on H and S, respectively (*contours*). H and S are on circular orbits around the centre of mass O of the two-body system. The First Lagrangian Point L_1 separates the Roche lobes of the host and the satellite. P is a test particle within the subhalo. The figure is for a satellite-to-host mass ratio of $M_\text{s}/M_\text{h} = 0.1$. It assumes that both the host halo and the subhalo are described by an NFW profile with $c = 8$ and that the subhalo lies at the virial radius of the host halo, which is used to scale the coordinates (so that $\overline{\text{HS}} = 1$). The figure is shown in a reference frame centred in O and corotating with the two-body system. A circular orbit around S through L_1 (*larger green dashed circle*) lies outside the Roche lobe of the satellite. A particle on this orbit is tidally stripped. The real value of r_t corresponds to the radius of the *smaller red dashed circle*, i.e. the largest circle centred on S to be entirely contained in the Roche lobe of the satellite.

be explored at the end of this appendix with a numerical experiment. By using (iii) and (iv), equation (A3) becomes $\dot{\boldsymbol{v}} = -\nabla\Phi_\text{eff}$, where

$$\Phi_\text{eff} = \Phi_\text{h} + \Phi_\text{s} - \frac{1}{2}\Omega^2\,\overline{\text{OP}}^2. \tag{A4}$$

Fig. A1 shows the equipotential contours for Φ_eff in a particular case used for illustrative purposes. In this particular case, $M_\text{s}/M_\text{h} = 0.1$, $\overline{\text{HS}} = R_\text{vir}$, and both the host and the satellite are described by an NFW profile with concentration $c = 8$ (R_vir is the host virial radius). The equipotential contours show a saddle point L_1 on the segment HS, which is the First Lagrangian Point L_1, which separates the Roche lobes of the host and the satellite. The particle P is tidally stripped if its orbit spills outside the Roche lobe of the satellite. In the next five paragraphs, we shall calculate $r_\text{t} \sim \text{S}\overline{L}_1$ in the limit that: (v) $M_\text{s} \ll M_\text{h}$, so that O → H, and (vi) $\overline{\text{SP}} \ll \overline{\text{HS}}$, so that we can expand $\boldsymbol{a}_\text{t}$ in powers of r/R (Jacobi limit). However, it is important to understand that, even if $\text{S}\overline{L}_1$ could be computed exactly, it would still provide an approximate estimate for r_t because a circle of radius $\text{S}\overline{L}_1$ and centre S lies outside the Roche lobe of the satellite. The implication is that the real tidal radius is $r_\text{t} < \text{S}\overline{L}_1$.

We can write $\vec{\text{OP}} = \boldsymbol{R} + \boldsymbol{r}$ in the limit that O → H (assumption v).[12] If the mass profile $M_\text{h}(R)$ of the host is spherically symmetric (assumption vii), equation (A3) can be rewritten as (Gonzalez-Casado et al. 1994)

$$\boldsymbol{a}_\text{t} = \left[-\frac{G\,M_\text{h}(|\boldsymbol{R} + \boldsymbol{r}|)}{|\boldsymbol{R} + \boldsymbol{r}|^3} + \frac{G\,M_\text{h}(R)}{R^3}\right](\boldsymbol{R} + \boldsymbol{r}), \tag{A5}$$

[12] Throughout this appendix, lowercase letters refer to distances from S and uppercase letters to distances from H.

where we have used equation (A1) in the limit $M_s \ll M_h$ to eliminate Ω^2 and we have neglected the Coriolis term because of assumption (*iv*).

Expanding a_t in a Taylor series to first order, we derive

$$a_t = \frac{G\,M_h(R)}{R^3}\left[-3 + \frac{R}{M_h(R)}\frac{dM_h(R)}{dR}\right]\frac{r}{R}\cos\theta\,R \qquad (A6)$$

By introducing the slope of the mean-density profile of the host

$$\alpha = \frac{d\ln\overline{\rho}}{d\ln r} = \frac{R}{M_h(R)}\frac{dM_h(R)}{dR} - 3\,, \qquad (A7)$$

Equation (A6) can be rewritten in the simpler form

$$a_t = \alpha\,\frac{G\,M_h(R)}{R^3}r\cos\theta\,\frac{R}{R}\,. \qquad (A8)$$

The term $\alpha\cos\theta\,R$ in equation (A8) is directed as $-\cos\theta\,R$ ($\alpha < 0$ because the density decreases with radius). In a system of polar coordinates (r, θ) centred on S,

$$\left(\alpha\cos\theta\,\frac{R}{R}\right)_r = |\alpha|\cos^2\theta, \qquad (A9)$$

$$\left(\alpha\cos\theta\,\frac{R}{R}\right)_\theta = |\alpha|\cos\theta\sin\theta, \qquad (A10)$$

where α appears in absolute value on the right-hand side because R points from the central system to the satellite (and thus $-R$ points like r when $\theta = 0$). Equation (A9) shows that the radial component of a_t is always positive, while equation (A10) shows that the azimuthal component has zero average. Therefore, the maximum acceleration is:

$$\langle a_{t,\mathrm{max}}\rangle = |\alpha|\frac{GM_h(R)}{R^3}r. \qquad (A11)$$

And averaging over all θ gives a mean outward acceleration of:

$$\langle a_t\rangle = \frac{|\alpha|}{2}\frac{GM_h(R)}{R^3}r. \qquad (A12)$$

We compute r_t by finding the point of the segment HS for which $a_t = -\nabla\Phi_s = GM_s(r)/r^2$, where $M_s(s)$ is the mass profile of the satellite, for which we assume spherical symmetry (assumption vii). Equation (A8) gives (Dekel, Devor & Hetzroni 2003)

$$\frac{G\,M_s(r_t)}{r_t^3} = |\alpha|\frac{G\,M_h(R)}{R^3}, \qquad (A13)$$

from which r_t can be computed numerically. There is no factor of two dividing $|\alpha|$ in equation (A13) because this equation is for $\theta = 0$ and not the result of an average ($\widehat{\mathrm{HSL}_1} = 0$).

The α parameter determines the strength of tidal stripping. The higher its absolute value, the lower the tidal radius r_t. For a DM halo described by the NFW profile, α decreases from $\alpha = -1$ at $R = 0$ to $-2.5 < \alpha < -2.2$ at $R = R \sim R_\mathrm{vir}$ (Fig. A2, black curve). The upper and lower limits for $\alpha(R_\mathrm{vir})$ corresponds to $c = 4$ and 12, respectively. The limit for $R \to \infty$, $\alpha = -3$, corresponds to the classical Jacobi limit for a point mass, whose average density over a sphere of radius R decreases as R^{-3}. The presence of a luminous galaxy at the centre of the halo causes the decrease of α to be much more rapid at first but then much slower because the limit at infinity has to be the same (Fig. A2, red curve). The value of α in the nearly flat part of the curve depends not only on c but also on the baryon-to-DM mass ratio and the baryon scalelength with respect to that of the DM. Despite these uncertainties, a value $-3 < \alpha < -2$ was to be expected, because the flatness of the rotation curves of spiral galaxies in their outer parts imply $\alpha \sim -2$ on the scale of the optical radius.
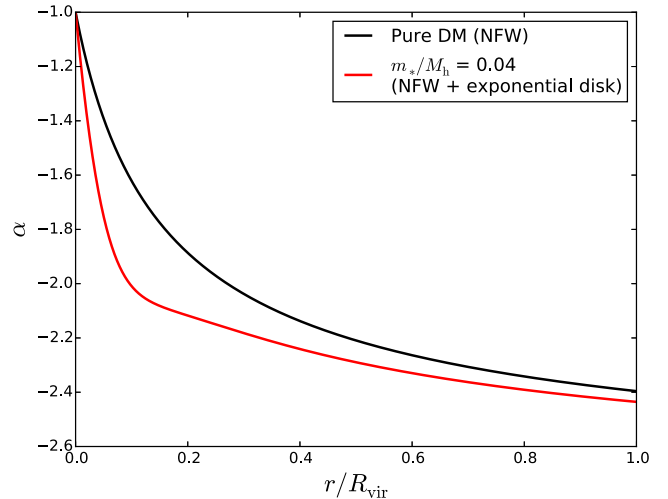

Figure A2. The logarithmic slope α of the mean-density profile (equation A7) as a function of the spherically averaged radial coordinate in units of the virial radius, for an NFW halo with $c = 8$ (*black solid line*) and the same halo with an additional exponential disc of mass equal to 4 per cent the total mass of the system (*red solid line*). Note that $m_*/M_h = 0.04$ is the maximum stellar-to-halo mass ratio allowed by AM. The exponential disc is assumed to have a scalelength $\lambda R_\mathrm{vir}/2 \sim 0.025 R_\mathrm{vir}$, where R_vir is the halo's virial radius.

To test the accuracy of equation (A13) in recovering the correct value of r_t, we have performed a numerical experiment, in which we start from the pure DM configuration in Fig. A1, we set up a range of initial conditions (r, v) for the test particle and we integrate their orbits to find under what conditions the particles escape from the satellite. The experiment retains the assumptions that the satellite is on a circular orbit (i), and that the mass distributions of the host and the satellite are spherically symmetric (vii), but allow us to relax the other five assumptions (the non-sphericity of discs has a small effect on the total gravitational potential of DM plus baryons).

We compute the position of L$_1$ by solving equation (A13) for $\alpha(R_\mathrm{vir}; c = 8) = -2.4$ and compare this result, derived from assumptions (v) and (vi), to the real position of L$_1$ in our configuration. Then, we consider four test particles with the same initial position (they all start at L$_1$ computed with equation A13), but with different initial velocities. Three of the four particles start on a circular orbit with $v_\mathrm{orb}^2 = GM_s(r_t)/r_t$, where v_orb is the orbital speed of P around S in an inertial frame. One corotates with Ω, another counter rotates and the orbital plane of the third one is orthogonal to that of the binary (the third case allows us to relax assumption iii). The fourth particle starts with zero speed at the apocentre of a purely radial orbit. All four are stripped from the satellite in less than an orbital time $(2\pi/\Omega)$, the duration of the numerical experiment.

We then progressively lower the value of α until the particles start so close to S that they are all able to remain within the satellite. Both $\alpha = -2.4$ (the value obtained from equation A7) and $\alpha = -3$ (the classical Jacobi limit) correspond to initial conditions for P in between the smaller and the larger dashed circle in Fig. A1. The α below which stripping is prevented depends on the initial condition for v, but the dependence is not strong. Hence, assumptions (iii) and (iv) are likely to have a minor effect on the value of r_t. Assuming $\alpha = -2.4$ overestimates r_t even in a pure DM configuration. So does $\alpha = -3$, but only slightly. The results of a numerical experiment in which P is positioned on the inner dashed circle of Fig. A1 are qualitatively similar to those shown in the right-hand panel of
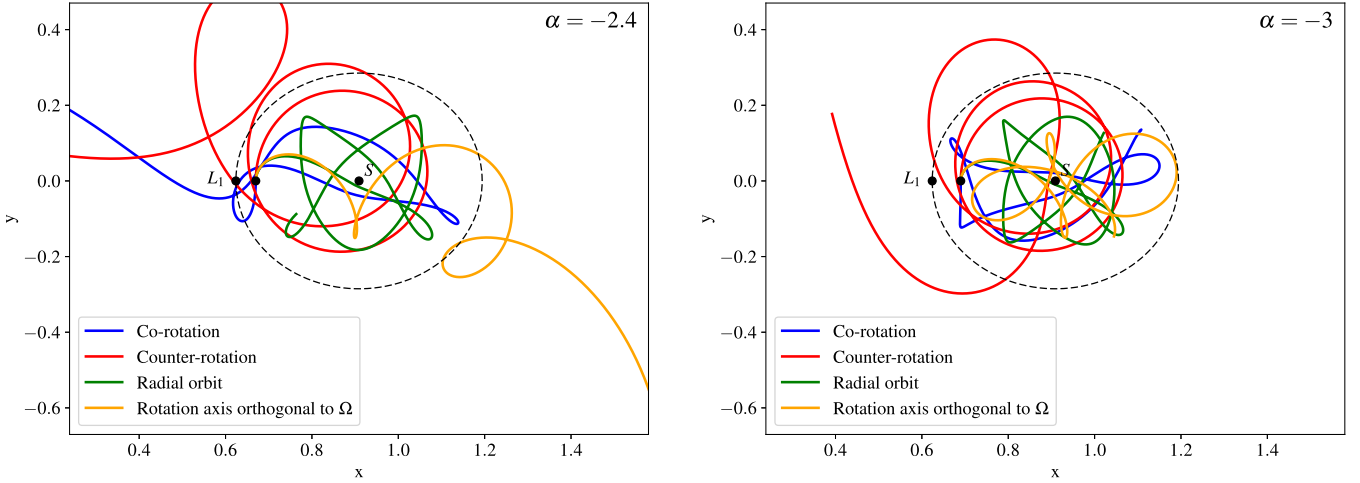
Figure A3. Orbits of four test particles during one revolution of S around O. All four test particles start at the position P of First Lagrangian Point computed with equation (A13), as opposed to the true First Lagrangian Point L_1, but their initial conditions for v are not the same: a particle corotates with Ω, another counter rotates with respect to Ω, a third one rotates on a plane orthogonal to the orbital plane of the two-body system and the fourth one starts on a purely radial orbit. The left- and the right-hand panels correspond to $\alpha = -2.4$, that is, $\alpha(R_{vir})$ for $c = 8$ and $\alpha = -3$, respectively. In the first case, all the particles on circular orbits are stripped from the satellite. Only the particle on a radial orbit (*green curve*) is retained. In the second case, only the counter-rotating particle (*red curve*) escapes from the satellite. All others are retained. The units on the *x*- and the *y*-axes are host-halo virial radii.

Fig. A3. The only difference is that the red curve makes many more orbits around S before it escapes from the satellite.

The numerical test in Fig. A3 illustrates the limitations of our analytic approach because it shows equation (A13) with $\alpha = -2.4$ from equation (A7) overestimates the real tidal radius (particles on circular orbits tend to be stripped even if they are at $r < r_t$). However, the right-hand panel of Fig. A3 shows that r_t is approximately recovered for a higher value of $|\alpha|$ corresponding to $\alpha = -3$. In principle, this value could depend on the distance between the satellite and the central, but Fig. A2 shows that, for $r/R_{vir} > 0.1$, α should be fairly independent of radius, especially in presence of baryons (red curve).

APPENDIX B: ORBITAL TIMES

Orbital times cannot simply be computed with energy conservation, because, as the halo increases in mass, its gravitational potential is not stationary. We have integrated the equation of motion $\ddot{r} = -\Phi'(r)\, r/r$ in a non-stationary NFW potential

$$\Phi(r, t) = -V_{vir}^2 \frac{\ln(1 + c\, r/r_{vir})}{f(c)\, r/r_{vir}} \, , \tag{B1}$$

where $f(c) = \ln(1 + c) - c/(1 + c)$ (Cole & Lacey 1996). We computed orbits by assuming that the potential varies in time as the median evolution of cosmological haloes. We ran the potential well growth historie code of van den Bosch et al. (2014) with the same cosmological parameters as used in our cosmological *N*-body simulation, saving the median halo mass and concentration as a function of redshift and lookback time. Our method is similar to that of van den Bosch et al. (2016). We considered a variety of initial apocentres, and both radial orbits, and orbits with initial apocentric velocity equal to 0.45 times the circular velocity at the apocentric radius. We measured the time from halo entry (first passage inside the evolving halo virial radius) to pericentre, to virial radius on the way out, and to the second apocentre. We fit the initial redshifts to reach these 3 radii at $z = 0$.

Fig. B1 displays the lookback times and corresponding redshifts for a galaxy to reach the pericentre (blue), virial radius on the way
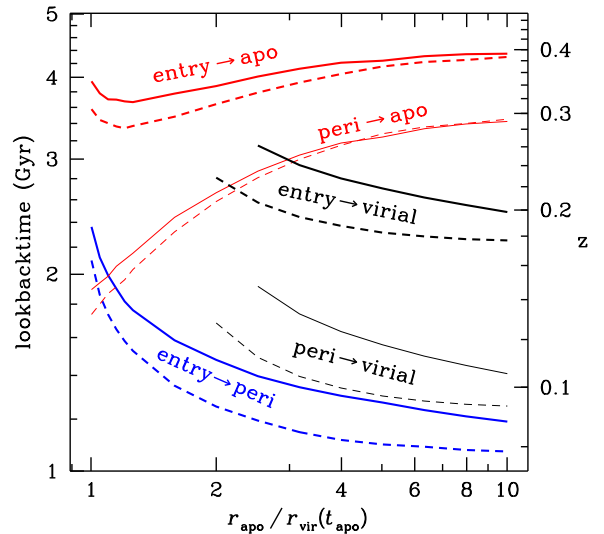
Figure B1. Lookback times and corresponding redshifts for galaxies to be at different locations at $z = 0$: pericentre (*blue*), virial radius on the way outwards (*black*) and apocentre (*red*), for radial orbits (*dashed*) and orbits of typical elongations (*solid*), starting the clock at cluster entry (*thick*) or at pericentre (*thin*).

out (black) and apocentre (red) at $z = 0$. While the dashed lines indicate the orbital times for radial orbits, the solid lines represent orbits of typical apocentre/pericentre ratios of 5 (Ghigna et al. 1998), where we used an apocentric (tangential) velocity of 0.45 times the circular velocity at that radius. The $z = 0$ halo mass is $10^{14}\, M_\odot$, but the results vary fairly little with the halo mass (except that the virial radius is harder to reach on the way out for more massive haloes). The times to reach the pericentre and the virial radius decrease for increasing initial apocentric radius (in units of the initial R_{vir}), because galaxies starting at large radii travel faster through the halo. On the other hand, the time to reach the second apocentre increases for increasing first apocentric radius (in units of the initial virial radius), because there is more distance to travel outwards (e.g. from

APPENDIX C: FITTING PARAMETERS FOR THE SMF AND THE SMHM RELATION

In this appendix, we explain how we use the data points of Yang et al. (2012) for the local Universe and Muzzin et al. (2013) at $z > 0.2$ to construct the SMF $n_*(m_*, z)$ that we use for AM.

We fit the data points in each redshift bin (centred on z_i) with a double-power-law function of the form:

$$n_i(m_*) = \frac{1}{N_i}\left[\left(\frac{m_*}{m_i}\right)^{\alpha_i} + \left(\frac{m_*}{m_i}\right)^{\beta_i}\right]^{-1}, \quad (C1)$$

where N_i, m_i, α_i and β_i are free parameters.

After having determined N_i, m_i, α_i and β_i for each z_i, we fit the evolution of N, m_*^0, α and β with z by assuming the linear dependencies:

$$\log N(z) = n_1 z + n_0$$
$$\log m_*^0(z) = x_1 z + x_0$$
$$\alpha(z) = \alpha_1 z + \alpha_0$$
$$\beta(z) = \beta_1 z + \beta_0 \quad (C2)$$

Table C1 gives the best-fitting values for the fitting parameters n_0, x_0, α_0, β_0 and n_1, x_1, α_1, β_1. The quality of the fit is shown on

Table C1. Best-fitting parameters characterizing the SMF at all redshift smaller than 2.5. The SMF is fitted at a given z by equation (C1) with four parameters which are assumed to be linear function of z (see equation C2).

n_0	x_0	α_0	β_0
2.39	11.12	0.228	3.07
n_1	x_1	α_1	β_1
0.501	4.6×10^{-3}	0.133	0.367
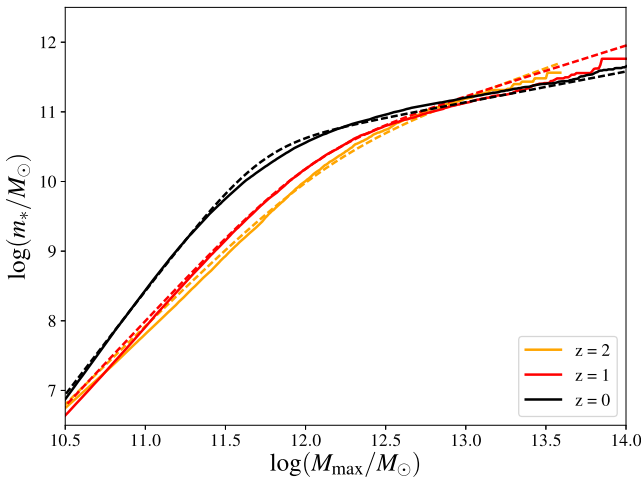
Figure C1. Comparison between the fitted stellar mass (m_*)–halo mass (M_{max}) relation (dotted lines) and the original relation used in this work (solid line) at $z = 0$, 1 and 2. The only purpose of this fitting is to facilitate the comparison with future AM work. The best-fitting parameters are given in Table C2.

Table C2. Best-fitting parameters characterizing the SMHM relation at all redshift smaller than 2.5. The SMHM relation is fitted at a given z by equation (C3) with four parameters which are assumed to be linear function of $z/(1 + z)$ (see equation C4).

m_0	y_0	γ_0	η_0
10.57	11.69	3.04	0.417
m_1	y_1	γ_1	η_1
−0.085	0.685	−1.16	0.607

Fig. 5. The parameters in Table C1 specify the SMF that we use for the AM.

To ease the comparison with future work, we apply the same fitting procedure to the SMHM relation. Following Moster et al. (2013), we obtain a good fit to the relation from AM (Fig. C1) for a double-power-law function of the form

$$m_*(m_{\mathrm{h}}, z) = M(z)\left[\left(\frac{m_{\mathrm{h}}}{m_{\mathrm{h}}^0(z)}\right)^{-\gamma(z)} + \left(\frac{m_{\mathrm{h}}}{m_{\mathrm{h}}^0(z)}\right)^{-\eta(z)}\right]^{-1}, \quad (C3)$$

where:

$$\log M(z) = m_1 \frac{z}{1+z} + m_0,$$
$$\log m_{\mathrm{h}}^0(z) = y_1 \frac{z}{1+z} + y_0,$$
$$\gamma(z) = \gamma_1 \frac{z}{1+z} + \gamma_0,$$
$$\eta(z) = \eta_1 \frac{z}{1+z} + \eta_0. \quad (C4)$$

Table C2 gives the best-fitting value for the fit parameters m_0, y_0, γ_0, η_0 and m_1, y_1, γ_1, η_1. We stress that the fitting formula in equation (C3) and the parameters in Table C2 are used nowhere in our analysis. They have been inserted purely to ease comparison with our work.

This paper has been typeset from a TeX/LaTeX file prepared by the author.